DESIGNING NEW NETWORK ADAPTATION AND ATM ADAPTATION LAYERS FOR INTERACTIVE MULTIMEDIA APPLICATIONS

THÈSE N^{O} 1782 (1998)

PRÉSENTÉE À LA SECTION DE SYSTÈMES DE COMMUNICATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES TECHNIQUES

PAR

Francisco Javier GARCIA ADANEZ

Ingénieur électricien diplômé EPF
de nationalité espagnole

acceptée sur proposition du jury:

Prof. J.-P. Hubaux, directeur de thèse
Prof. E. Biersack, corapporteur
Prof. G. Karlsson, corapporteur
Dr T. Stock, corapporteur

Lausanne, EPFL
1998

Life is just too complicated to be smart all the time

Scott Adams

Acknowledgments

My deepest thanks go to my advisor Professor Jean-Pierre Hubaux who accepted me in his laboratory, supported me all along these years and definitely contributed to the success of this work.

I would like to express a deep gratefulness to Dr. Thomas Stock for his invaluable contribution to this work. His friendliness, advice and support encouraged me all over this work.

I am deeply grateful to Dr. Reto Gruenenfelder who originated the INCOMS project that served as a basis for this thesis.

I owe a lot and I am particulary grateful to Olivier Verscheure with whom I had the opportunity to discuss and share ideas. He undoubtedly helped to the success of this work with his contribution to several publications and his knowledge of vision science.

I express a special thank you to the members of the multimedia group: Pascal Frossard for the discussions and the simulation work; Raffaele Noro with whom I had the privilege of sharing the office and learning some Italian. His comments and valuable suggestions have been of great help during the last two years; Dr. Maher Hamdi for his support and proofreading of the last chapters of this thesis.

Nothing could have been done without the essential help and support of Bruno Dufresne who provided the computing environment. His availability and good mood have largely contributed to the completion of this work.

I have a special gratitude towards all the members of the TCOM group, and particularly towards Cesar Gallego and Laurent Jaussi for their helpful collaboration and fruitful dicussions. I also want to specially thank Danielle Alvarez for making things a lot easier.

I also would like to thank several persons that have contributed in different ways to this work: Dr. Gunnar Karlsson, Dr. Christian van den Branden Lambrecht, Dr. Jean-Paul Gaspoz and Dr. Sam Manthorpe.

Finally, my gratitude goes to my wife Fabienne who patiently encouraged me during all the years spent at EPFL. This work owes a lot to her presence, love and support.

Abstract

Multimedia services, audiovisual applications composed of a combination of discrete and continuous data streams, will be a major part of the traffic flowing in the next generation of high speed networks. The cornerstones for multimedia are Asynchronous Transfer Mode (ATM) foreseen as the technology for the future Broadband Integrated Services Digital Network (B-ISDN) and audio and video compression algorithms such as MPEG-2 that reduce applications bandwidth requirements. Powerful desktop computers available today can integrate seamlessly the network access and the applications and thus bring the new multimedia services to home and business users. Among these services, those based on multipoint capabilities are expected to play a major role.

Interactive multimedia applications unlike traditional data transfer applications have stringent simultaneous requirements in terms of loss and delay jitter due to the nature of audiovisual information. In addition, such stream-based applications deliver data at a variable rate, in particular if a constant quality is required.

ATM, is able to integrate traffic of different nature within a single network creating interactions of different types that translate into delay jitter and loss. Traditional protocol layers do not have the appropriate mechanisms to provide the required network quality of service (QoS) for such interactive variable bit rate (VBR) multimedia multipoint applications. This lack of functionalities calls for the design of protocol layers with the appropriate functions to handle the stringent requirements of multimedia.

This thesis contributes to the solution of this problem by proposing *new Network Adaptation and ATM Adaptation Layers* for interactive VBR multimedia multipoint services.

The foundations to build these new multimedia protocol layers are twofold; the requirements of real-time multimedia applications and the nature of compressed audiovisual data.

On this basis, we present a set of design principles we consider as mandatory for a generic *Multimedia AAL* capable of handling interactive VBR multimedia applications in point-to-point as well as multicast environments. These design principles are then used as a foundation to derive a first set of functions for the MAAL, namely; cell loss detection via sequence numbering, packet delineation, dummy cell insertion and cell loss correction via RSE FEC techniques.

The proposed functions, partly based on some theoretical studies, are implemented and evaluated in a simulated environment. Performances are evaluated from the network point of view using classic metrics such as cell and packet loss. We also study the behavior of the cell loss process in order to evaluate the efficiency to be expected from the proposed cell loss correction method. We also discuss the difficulties to map network QoS parameters to user QoS parameters for multimedia applications and especially for video information. In order to present a complete performance evaluation that is also meaningful to the end-user, we make use of the MPQM metric to map the obtained network performance results to a user level. We evaluate the impact that cell loss has onto video and also the improvements achieved with the MAAL.

All performance results are compared to an equivalent implementation based on AAL5, as specified by the current ITU-T and ATM Forum standards.

An AAL has to be by definition generic. But to fully exploit the functionalities of the AAL layer, it is necessary to have a protocol layer that will efficiently interface the network and the applications. This role is devoted to the Network Adaptation Layer.

The network adaptation layer (NAL) we propose, aims at efficiently interface the applications to the underlying network to achieve a reliable but low overhead transmission of

video streams. Since this requires an *a priori* knowledge of the information structure to be transmitted, we propose the NAL to be codec specific.

The NAL targets interactive multimedia applications. These applications share a set of common requirements independent of the encoding scheme used. This calls for the definition of a set of design principles that should be shared by any NAL even if the implementation of the functions themselves is codec specific. On the basis of the design principles, we derive the common functions that NALs have to perform which are mainly two; the segmentation and reassembly of data packets and the selective data protection.

On this basis, we develop an MPEG-2 specific NAL. It provides a perceptual syntactic information protection, the PSIP, which results in an intelligent and minimum overhead protection of video information. The PSIP takes advantage of the hierarchical organization of the compressed video data, common to the majority of the compression algorithms, to perform a selective data protection based on the perceptual relevance of the syntactic information.

The transmission over the combined NAL-MAAL layers shows significant improvement in terms of CLR and perceptual quality compared to equivalent transmissions over AAL5 with the same overhead.

The usage of the MPQM as a performance metric, which is one of the main contributions of this thesis, leads to a very interesting observation. The experimental results show that for unexpectedly high CLRs, the average perceptual quality remains close to the original value. The economical potential of such an observation is very important. Given that the data flows are VBR, it is possible to improve network utilization by means of statistical multiplexing. It is therefore possible to reduce the cost per communication by increasing the number of connections with a minimal loss in quality.

This conclusion could not have been derived without the combined usage of perceptual and network QoS metrics, which have been able to unveil the economic potential of perceptually protected streams.

The proposed concepts are finally tested in a real environment where a proof-of-concept implementation of the MAAL has shown a behavior close to the simulated results therefore validating the proposed multimedia protocol layers.

Version Abrégée

Les applications multimédia qui combinent différents flux de données audiovisuelles occuperont une place importante dans la prochaine génération de réseaux à haut débit. Les composantes clé du multimédia sont le mode de transfert asynchrone (ATM) considéré comme la technologie du futur réseau numerique à integration de services à large bande (B-ISDN) ainsi que les algorithmes de compression de données audiovisuelles comme MPEG-2 qui réduisent les besoins en largeur de bande des applications. De plus, la disponibilité d'ordinateurs capables d'intégrer de façon transparente l'accès au reseau est en train de généraliser les applications multimédia tant dans les entreprises que pour les loisirs. Parmi ces applications, celles basées sur le multipoint seront appelées à jouer un rôle majeur.

Contrairement aux applications de transfert de données traditionnelles, les applications multimédia interactives ont des fortes contraintes en termes de pertes et de délais de par la nature même des informations audiovisuelles. En outre, ces applications delivrent les données à un débit variable, en particulier si une qualité vidéo constante est requise.

ATM permet d'integrer différents types de trafic dans un même réseau ce qui crée des interactions qui se traduisent par des délais et des pertes. Les protocoles traditionnels ne possedent pas les mécanismes appropriés pour offrir la qualité de service (QoS) réseau requise par les applications multimédia interactives à débit variable. Ce manque de fonctionalités requiert le design de nouvelles couches de protocoles capables de gérer les contraintes propres à ces applications.

Cette thèse contribue à résoudre ce problème en proposant de nouvelles couches d'adaptation réseau et ATM pour des services multipoint multimédia interactifs à débit variable.

Les nouvelles couches de protocole proposées sont développées d'après les besoins des applications multimédia temps-réel et la nature des données audiovisuelles compressées.

Sur cette base, nous présentons une série de principes estimés comme nécessaires pour une AAL multimédia génerique capable de gérer des applications interactives à débit variable tant dans des environnements point-à-point que multipoint. Cette série de principes de conception sont alors utilisés pour développer les fonctions de la MAAL, à savoir, détection de cellules perdues par un numéro de séquence, segmentation des paquets, insertion de cellules vides et correction d'erreurs via des techniques FEC basées sur des codes de type RSE.

Les fonctions proposées, en partie basées sur des études théoriques, sont implementées dans un environnement simulé. Les performances sont evaluées d'un point de vue réseau en utilisant des métriques classiques comme les pertes de cellules et de paquets. On étudie aussi le comportement des processus de pertes de façon à évaluer l'efficacité du mécanisme de correction de pertes proposé.

Ce travail aborde aussi le problème du “mapping” entre les paramètres de qualité de service réseau et utilisateur en particulier en ce qui concerne la vidéo. De façon à présenter une évaluation de performances complète qui ait aussi un sens pour l'utilisateur, nous utilisons une mesure de qualité vidéo perceptive, MPQM. Celle-ci permet de traduire les résultats obtenus au niveau du réseau vers le niveau utilisateur. Nous évaluons aussi l'impact que les pertes de cellules ont sur la vidéo et les améliorations obtenues avec la couche MAAL. Les résultats sont comparés avec une implementation equivalente basée sur AAL5, comme specifié par les standards actuels de l'ITU-T et de l'ATM Forum.

Une AAL doit par définition etre générique. Mais pour exploiter au maximum ses fonctionalités, il est nécessaire d'avoir une couche de protocole qui interfacera de façon efficace le réseau et les applications. Ce rôle est consacré à la couche d'adaptation réseau (NAL).

Cette couche d'adaptation a pour but d'interfacer de façon efficace les applications au réseau pour obtenir une transmission fiable, bien qu'avec un faible supplément d'information, de flux audiovisuels. Ceci nécessite une connaissance *à priori* de la structure de l'information à transmettre. Par conséquent, la couche NAL doit être spécifique à un codeur.

Les algorithmes de compression utilisés dans ces applications partagent une série de besoins communs indépendants du schéma d'encodage utilisé. Il est donc nécessaire de définir une série de principes de fonctionnement qui devraient être partagés par toutes les NALs, même si l'implémentation des fonctions elles-mêmes dépend de l'encodeur. Sur cette base nous définissons les fonctions propres aux couches NAL, à savoir: segmentation et reassemblage des paquets et protection selective des données.

A partir de ces fonctions nous avons développé une couche NAL spécifique au codeur MPEG-2. Elle offre une protection syntaxique perceptive de l'information, le PSIP, qui réalise une protection intelligente des flux vidéo avec un surplus d'information minimum. Le PSIP profite de l'organisation structurée et hierarchique des données compressées pour effectuer une protection selective des informations basée sur l'importance perceptive des données syntaxiques.

La transmission de flux audiovisuels sur les deux couches de protocole NAL-MAAL montre des améliorations significatives au niveau du CLR ainsi qu'au niveau perceptif comparé à une transmission équivalente sur AAL5.

L'utilisation de MPQM comme mesure de performance, ce qui constitue une des contributions majeures de cette thèse, mène à une observation très importante. Les résultats expérimentaux montrent que pour des valeurs très élevées du CLR, la qualité perceptive moyenne reste très proche de la valeur originale. Le potentiel économique d'un tel phénomène est très important. Etant donné que les flux de données sont de type VBR, il est possible d'augmenter l'utilisation du réseau au moyen du multiplexage statistique. Il est ainsi possible de réduire le coût par communication en augmentant le nombre de connexions avec une reduction minimale de la qualité perceptive.

Cette conclusion n'aurait pas pu être obtenue sans l'utilisation combinée des mesures de qualité perceptive et réseau, qui ont permis de dévoiler le potentiel économique des flux protégés de façon selective et perceptive.

Les concepts proposés sont finalement testés dans un environnement réel dans lequel une implementation de la MAAL démontre un comportement similaire à celui obtenu par simulation validant ainsi les couches de protocole multimédia proposeées.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Summary

Telecommunications are one of today's economic driving forces. Market deregulation and the strategical need for information have fueled the development of communication services and applications. Among these, interactive multimedia applications are gaining momentum fostered by web technology and reach now various markets. The development of corporate intranets opens the door to video conferencing, distributed cooperative work and other multimedia based services such as distant training and repairing. Residential access via technologies such as xDSL and cable TV is being pushed by the entertainment and the personal computer industries. The perspective of broadband access to the internet and a large spectrum of on demand services lets foresee mass markets still to be exploited. The creation of large companies resulting from the merging of communications and entertainment enterprises illustrates the strategic importance of business and residential customers.

The interest that such large markets generate have fostered the work of standardization bodies in order to offer the specifications necessary for a large scale deployment of network infrastructures able to carry such multimedia services.

Multimedia results from the combination of several enabler technologies issued by different technical communities. Integrating these technologies, originally developed without any particular vision of integration, has been the main work of the standards bodies. However, the resulting set of specifications for multimedia has been generally driven by a requirement of fast deployment instead of a true wish of integration of several already mature technologies.

This situation is illustrated by the efforts done by a plethora of existing and new standardization and specification organizations such as ITU, DAVIC, ATM Forum, IETF, ISO, ADSL Forum, etc...

Multimedia is at the intersection of traditionally separated industries as depicted in Fig. 1.1. This intersection reflects the nature of multimedia applications which integrate multiple media in a single application. The transmission of such applications requires a network capable of handling many different types of data. Since several years, the solution has been considered to be *Asynchronous Transfer Mode* (ATM). ATM is the network technology at the core of the Broadband Integrated Services Digital Network (B-ISDN) capable of achieving the level of integration required by multimedia applications.

However, a really integrated network will have to cope with very different requirements in terms of QoS, delay, delay jitter and data loss. Such an integration has proven to be difficult to achieve while providing adequate QoS to all users.

One of the problems is that audiovisual compression standards have barely been done

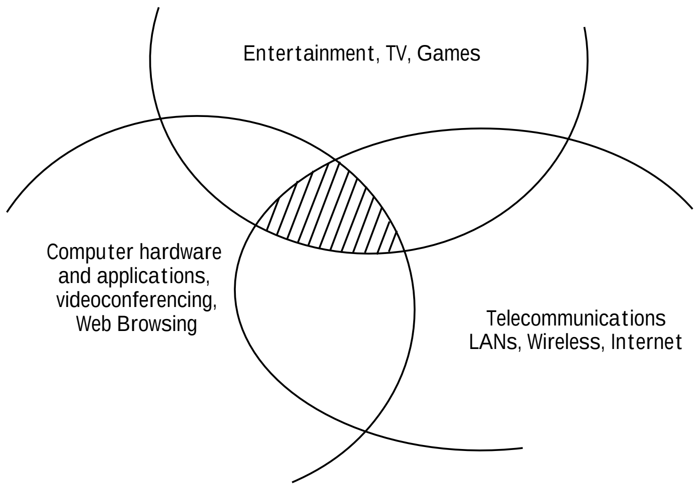


Figure 1.1: *Convergence of multimedia enabler industries.*

with transmission in mind. Simultaneously, broadband transmission specifications have not been developed with multimedia applications in mind. This gives us today protocol layers and in particular ATM Adaptation layers (AAL) not adapted to interactive multimedia applications requirements.

The problem to solve is how to tie together multimedia applications requirements and network functions. In particular if the following issues are considered:

- traffic characteristics in fully integrated networks may be different from what can be observed in today's networks
- audiovisual applications will generate variable data rates to deliver constant quality video and moreover will have very tight timing constraints if they are interactive
- from a networking standpoint, Constant Bit Rate (CBR) traffic is easier to handle than Variable Bit Rate (VBR)
- statistical multiplexing achieves a better network utilization and accommodates better VBR traffic
- there is no AAL adapted for such a large palette of emerging and future multimedia applications.

These obstacles in addition to new future services and applications not foreseen today calls for the development of new protocol layers suited for the specific characteristics of multimedia applications namely; highly structured, low redundancy and time constrained information.

The scope of this thesis is the design of protocol layers tailored to the requirements of interactive multimedia applications. This work proposes an ATM Adaptation Layer to provide low delay and reliable transmission as required by multimedia applications. An enhancement layer called Network Adaptation Layer (NAL) on top of the AAL is also proposed which takes full advantage of the error detection and correction functions provided by the AAL. The NAL provides a selective error protection mechanism based on the perceptual relevance of the data to be transmitted.

Ultimately, the target of this work is to reduce the cost of interactive multimedia communications by improving the usage of network resources.

1.2 Methodology Adopted in this Work

1.2.1 Understanding Multimedia Requirements

To derive appropriate mechanisms for the transmission of interactive multimedia data, it is essential to understand the requirements of such applications. Unlike traditional data communications, multimedia directly involves the user. The implications of this human involvement have important consequences for the requirements and design of multimedia applications. From the user perspective, interactive applications must display the data in real-time, that is with as low delay as possible and synchronized. In addition, the quality of the displayed material has to be as good as possible. However, the masking phenomenon of the human visual system allows to tolerate some, yet limited, loss in the data. From the network point of view, there are two basic requirements: low delay and low loss.

This illustrates some of the major differences between multimedia and traditional data. The latter does not tolerate any loss but has much more relaxed constraints in terms of delay.

Another major difference between multimedia data and traditional data is the organization of the information. Compressed audiovisual streams are hierarchically structured. All data are *not equally important*. The separation into syntactic and semantic components requires specific transmission mechanisms able to treat differently both types of data in order to reduce the impact that lost information may have onto the displayed material. This is not the case for data communications where all data have the same degree of importance. In that sense, traditional protocol layers do not provide the required functions to efficiently convey multimedia data.

The understanding of such essential differences between traditional information and multimedia streams is the first step towards the design of protocol layers adapted to interactive multimedia application requirements and is covered in the first part of this thesis.

1.2.2 Designing Appropriate Network Mechanisms

Once the application and user requirements are understood it is possible to derive a set of essential functions multimedia protocols must have. These drive the design principles to be used as a foundation for the development of such multimedia protocol layers.

The application of these principles lead to two protocol layers: the MAAL and the NAL. The former is a generic light weight protocol layer capable of handling low delay, VBR data streams. The latter is a codec-specific protocol layer which integrates the concept of perceptual relevance of the data.

The general framework proposed in this thesis is built upon these two multimedia-oriented layers.

The test, verification and validation of the design is made via simulation and a prototype implementation. The testbeds, simulated and experimental, have been used with MPEG-2 based applications. The results may then be used to further refine and improve the stack performance.

Such an iterative approach allows to fine tune the protocol's functions in order to achieve the best possible performance.

1.3 Main Contributions

The main contributions that this thesis proposes are:

- **the design of a new ATM Adaptation Layer (AAL)** for multimedia communications. The AAL called MAAL covers the requirements of multimedia applications in terms of delay and loss. Perceptual quality being a key design feature, the protocol layer handles VBR streams, since VBR encoding is required to achieve near-constant quality video. The MAAL provides a sequence number and a Reed-Solomom erasure forward error correction (FEC) mechanism at the cell level. The error correction scheme does not require interleaving thus avoiding large delays

- **the design of a Network Adaptation Layer (NAL)**. The NAL's main functions derived in this thesis are *generic* and applicable to any type of audiovisual data. The main NAL design criterion is the *user perceived quality*. It provides a selective protection mechanism based on the perceptual relevance of the information which requires an a priori knowledge of the data. Therefore, the NAL must be codec specific. In this thesis, an MPEG-2 specific NAL has been developed. The MPEG-2 NAL minimizes network loss impact while generating only a small overhead

- **the introduction of psychophysics as a performance tool for network QoS**. Such perceptual quality metrics are *user-oriented metrics* that have been used for network performance assessment. This innovative approach gives and insight for QoS mapping between the user and the network. The application of such metrics shows the difficulty to map performance results between the user and the network

- **the design of a Perceptual Syntactic Information Protection (PSIP) mechanism** which takes into account the perceptual relevance of the data in terms of loss impact. The PSIP algorithm is a new technique that selectively protects the syntactic data in an audiovisual stream. The concept has been applied to MPEG-2 streams but is generic and applicable to any kind of structured information

- **the design and implementation of a proof-of-concept software prototype of the MAAL protocol layer**. The prototype demonstrates the advantages of the concept of lightweight protocols and confirms the results obtained by simulation. The prototype achieves data transfer rates of up to 6 Mbps on top of the board's API

- **the introduction of the concept of protocol applet or proclet**. The NAL functions being codec specific, the development of a NAL per codec may provoke a cluttering of layers in the protocol stack. The concept of proclet is a new concept of dynamically configurable protocols that could be downloaded according to the requirements of the application onto the receiver. Such mobile protocols must be simple. Proclets are driven by Java and Javascript technology.

1.4 Structure of the document

The first part of this thesis, studies in detail the requirements that interactive multimedia applications have, related to the user and the application. This study developed in Chap. 2 is completed by a review of the current network functions available for such interactive services. From this study, we derive a set of functions that the network has to provide to ensure a good *quality of service* to the user. The set of functions we consider as essential are then compared in Chap. 3 to the current state-of-the art networking technologies.

Part two of this thesis develops and tests new protocol layers for multimedia applications. In Chap. 4 we derive the functions and structure of the multimedia AAL. Assumptions are introduced that allows to perform some mathematical developments used for dimensioning. The MAAL proposed is then tested in Chap. 5 in a simulated environment. We introduce perceptual quality metrics as a network performance tool. The results show improvements over AAL5 in terms of network QoS and user perceived quality mainly due to the mechanisms adapted to multimedia streams. Chapter 6 develops the concept of network adaptation. The functions derived are used for the development of an MPEG-2 specific network adaptation layer. Further simulations use the perceptual syntactic information protection which selectively protects the data. The proposed layers achieve much better results than the transmission over AAL5 and show a potential for statistical multiplexing gain which lets foresee cost reductions for such services. A proof-of-concept prototype implementation of the MAAL is described in Chap. 7. The experimental results confirm the simulation results and prove the viability of the concept of a lightweight protocol layer for multimedia. Achievements and possible extensions are described in Chap. 8 which eventually concludes this dissertation.

Chapter 2

Multimedia Applications and the B-ISDN Paradigm

Until recent years, there have been distinct networks for different types of applications. Telephony, data and TV distribution services did not share network resources for decades. The user needed separate and specific devices to access each of the networks; the telephone set for the circuit switched telephone network, the computer via a modem or a Packet Assembler-Disassembler (PAD) for the data transmission network and the TV set for the cable or wireless TV distribution network.

Today, a single device, the computer, can display information by means of multiple media which include audio, video, graphics and still images in addition to the well known text medium. The fast increase of CPU power, data storage as well as a price reduction of essential hardware components such as Random Access Memory (RAM) made possible the introduction of personal computes into the mass market which fostered the development of multimedia applications. Simultaneously, the development of powerful image compression algorithms has reduced the space required to store video sequences and most important has drastically reduced the bandwidth requirements for video transmission which today enables the delivery of individual audiovisual services and multimedia applications in general over data and telecommunication networks.

However, this convergence of multiple media in a single terminal device requires transmission over a network capable of handling simultaneously data flows of very different nature. Such a generic network is called an *integrated network*. The development of this concept originates with the introduction of the Integrated Services Digital Network (ISDN) [1]. However, the low bandwidth and limited flexibility of ISDN kept the network integration at the concept state. The development of high speed networking technology and in particular ATM, introduced the bandwidth and the required flexibility to handle multiple types of data by means of the Quality of Service (QoS) concept. This gave a new impulse to the concept of integrated network and lead to the *Broadband Integrated Services Digital Network* (B-ISDN) paradigm. It improved the capability of transmitting large amounts of information in real-time giving the possibility of interaction between multiple distributed users. Several international consortia such as the Digital Audio VIsual Council (DAVIC), the ATM Forum and the Multimedia Computer Forum (MMCF) are now specifying the ways of transmitting real-time multimedia applications over such high speed networks.

Another important force driving the fast development of multimedia has been the introduction of the *World Wide Web*. In spite of the existence of the internet for more than 20 years, the introduction of a simple and widely available generic multimedia interface has brought the computer network to the home opening a broad market which today is still difficult to evaluate. However, the extraordinary traffic generated by the millions of users in the internet compared to the relatively scarce resources and the use of unsuited protocols for continuous data streams make it very difficult to obtain a good quality of service for such applications. Efforts are currently under way to improve this situation. A resource reservation protocol called RSVP [2] provides resource allocation in IP based networks. The next generation of the Internet Protocol IPv6 [3] will provide means to allow guaranteed QoS in combination with RSVP. Also, new transport protocols such as the Real-time Transport Protocol (RTP) [4] have been developed to offset the limitations of the current set of protocols in use in the internet.

Such combination of multimedia applications and high speed networks provides large possibilities to develop a wide palette of new applications in very diverse domains which today cover entertainment (Video on Demand -VoD-, Karaoke on demand), business (collaborative work, telecommuting), education (interactive distance learning, teleteaching) and health care (telemedicine). All these applications make extensive use of audiovisual features. However, the nature of the data generated by multimedia applications is different from the one generated by legacy file transfer or terminal applications. In computer networks, traditional file transfers tried to be done as fast as possible, but there were no timing constraints. The only fundamental requirement for such applications was the correctness of the transmitted data.

Most of the audiovisual data is of a *continuous* nature. A continuous data stream has a temporal dimension which entails a *time-dependence* property. As such, it has very stringent constraints in terms of delay and delay jitter. This has been referred as *timely information* [5]. This means that data arriving beyond a certain point in time is not valid anymore and is considered as lost by the application. In fact, even if these streams are referred to as continuous, they are discrete. It is the inertia of the human senses that gives the impression of continuity. It is therefore clear that the characteristics of multimedia applications are strongly linked to the user and to the user's perception of the displayed data. Multimedia applications add a temporal dimension that data applications do not have. The user has basic perceptual requirements that are related to the quality and continuity of the image and sound, the synchronization of the correlated flows and the reaction time to interaction. To achieve these requirements, today's networks have to satisfy several constraints in terms of delay, delay jitter and errors.

The rest of this chapter describes in detail what the user requirements are, the consequences that these impose onto networking concepts and the state of the network technology for B-ISDN.

2.1 Requirements of Multimedia Applications

2.1.1 User Requirements

The user requirements are based on perception. The most important *perceptual channels* of information for humans with respect to current multimedia applications are the ear and the eye. However, if taken separately, each one has a certain degree of error tolerance. Humans are rather intolerant to audio errors. The compressed audio signal does not require many bits to be represented which makes it very sensitive to errors of several consecutive bits. Conversely, we are quite tolerant to visual errors. Uncompressed video data has simultaneous spatial and temporal redundancy which makes it relatively low sensitive to errors. It also has important spatial and temporal correlation. Compared to audio, however, the redundancy makes this correlation less important concerning information faults. Moreover, there are also other issues that could make errors significant or not (e.g. the focus of interest).

For non-continuous media, the situation is rather different. Text or still pictures, are almost not tolerant to errors. In general these pieces of information are transmitted as separate files and if there are errors, the applications cannot interpret, and therefore, cannot display them. Even if they are displayed these errors are rapidly perceived by the user because they last until the information is replaced, which can last between seconds to minutes.

Beyond the perception of isolated media, is the perception of combined media. The integration of different media is the characteristic of multimedia applications. This integration entails temporal relations between the different media objects. The respect of this timing dependency is called *intra-synchronization*. The importance of synchronization, to some extent, depends on the application. How the user perceives the loss of synchronization also depends on the media. The lip synchronization or *skew effect* between audio and associated video has been the object of several studies. There is no exact rule to formulate what the constraints are because the skew effect may depend on the scene contents. However, as a general rule $\pm$ 80 ms is considered as the threshold beyond which the user notices an out of synchronization effect for one-way video [6, 7].

Another factor which may influence our perception of the quality is the nature of the application. We distinguish two classes of real-time applications:

- non-interactive applications which have strict real-time requirements but can tolerate an initial delay in the playout of the data that will not be noticed by the end-user. This is the case of Video on Demand (VoD) applications. These applications require enough bandwidth and a bounded delay but low delay values are not strictly necessary. To absorb the initial delay, some data is buffered at the receiver. The advantage is that the buffer may be used for jitter absorption and flow synchronization and to some extent to perform error recovery
- interactive applications which include video conferencing and live interactive broadcasting. These applications deliver live audio and video and allow for some level of interaction. Also applications like near-VoD which do not deliver live data offer some degree of interactivity via VCR-like controls. These applications have the same bandwidth requirements as non-interactive applications but also require low end-to-end delay to provide the desired level of interaction.

In summary, the typical user requirements are low latency especially for interactive services, synchronization between the media, low loss on audio and a higher yet limited fault tolerance on video. It is clear that all these requirements have a strong influence on the applications and on the transmission requirements aspects that are developed in the next sections.

2.1.2 Application Requirements

Multimedia applications have to fulfill all the user requirements. Basically, applications have to deliver the different media objects in time, synchronized and with as few perceptive errors as possible. However, their performance will depend on the network's ability to fulfill the following requirements:

- deliver data in time (low delay and low delay variation)
- deliver multiple data flows with a minimum delay variation between the flows (synchronization)
- deliver the data with a minimal number of errors.

All these requirements and the resulting quality of service will therefore depend on the network's ability to provide a low loss, low delay and low delay jitter transmission. In return, the network performance will depend on the characteristics of the data that the sender's application delivers to the network.

2.1.2.1 Video Compression

One of the major developments that have enabled networked multimedia is the development of powerful *image compression* algorithms [8, 9]. Originally, video data has always been extremely bandwidth consuming as shown in Tbl. 2.1. Compression has reduced the bandwidth needs giving the possibility of transmitting video over computer or packet networks. However, this had important consequences on the video stream that the application delivers to the network.

Application	Spatial Resolution	Uncompressed Bit Rate (RGB)	Compressed Bit Rate
NTSC video	$720 \times 480 \times 29.97Hz$	168 Mbps	4-8 Mbps
PAL video	$720 \times 576 \times 25Hz$	199 Mbps	4-9 Mbps
Digital HDTV	$1920 \times 1080 \times 30Hz$	1493 Mbps	18-30 Mbps
Digital HDTV	$1280 \times 720 \times 60Hz$	1327 Mbps	18-30 Mbps
ISDN Video Telephony	$352 \times 288 \times 29.97Hz$ (CIF)	73 Mbps	64-1920 Kbps
PSTN Video Telephony	$176 \times 144 \times 29.97Hz$ (QCIF)	18 Mbps	10-30 Kbps
Two-channel stereo audio		1.4 Mbps	128- 384 Kbps
Five-channel stereo audio		3.5 Mbps	384-968 Kbps

Table 2.1: *Raw and compressed video rates for various applications.*

The principle of compression algorithms is to reduce the redundancy existing in video signals which is of two types: spatial (within a single frame) and temporal (between adjacent frames). Furthermore, compression algorithms can be divided into *lossless* and *lossy* algorithms depending on the type of coding used.

Entropy coding is lossless because it uses only statistical properties of the data and does not take into account the semantics of the information to compress (e.g. run length coding). Such lossless algorithms allow to recover entirely the original data from the compressed one, they are *reversible.* These techniques find use in applications such as data storage and medical imaging. Also, lossless techniques are generally symmetrical which means that compression and decompression are equally computationally complex. However, such algorithms remove less redundancy and therefore cannot achieve high compression ratios (generally up to 15:1). The compression algorithms achieve compression ratios high enough to make economically feasible the transmission of, at least, cable television (CATV) quality video over packet networks.

The compression algorithms using source coding are lossy. These algorithms take into account the semantics of the data and are therefore able to achieve higher compression rates. They remove subjectively redundant information. To still achieve a high quality video, the removed information has to be as little perceptually visible as possible. However this type of compression removes part of the original data which therefore cannot be recovered and is called lossy compression. MPEG [10] and H.261 [11] are examples of such algorithms. In addition, these algorithms use the temporal redundancy inherent to motion video. The original data cannot be recovered from the compressed one due to a more important redundancy reduction. This allows to achieve much higher compression ratios, in the order of 100:1 or 200:1. These algorithms are generally asymmetrical. Encoding is much more complex than decoding which has the advantage of making decompression possible in low end equipment. Such algorithms are generally used in applications where compression is performed once and decompression several times (e.g. Video on Demand). The majority of the available compression algorithms use a combination different techniques known as *hybrid coding*.

The most widely used compression algorithms are based on transform coding. In a transform coding system, pixels are grouped into blocks. A block of pixels is transformed into another domain to produce a set of transform coefficients which are then coded and transmitted.

The Discrete Cosine Transform (DCT) is today the most popular and widespread transform coding method used for several reasons; it provides good energy compaction and several fast algorithms exist for calculating the DCT of a block of pixels.

The DCT converts a block of pixels into a block of transform coefficients of the same dimensions. These coefficients represent the spatial frequency components of the original pixel block. Eq. 2.1 shows the two dimensional DCT of an $N \times N$ pixel block. $f(i,j)$ represents the pixel values and $F(u,v)$ the transform coefficients.

$$F(u,v) = \frac{2}{N} C(u)C(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2i+1)v\pi}{2N}\right), \tag{2.1}$$

where

$$C(x) = \begin{cases} \frac{1}{\sqrt{2}}, \; x = 0 \\ 1 \; otherwise. \end{cases} \tag{2.2}$$

Performing such transform does not produce any compression. It however, groups the larger value coefficients around the DC value since in a typical image, the blocks of pixels tend to have little high-frequency contents. Therefore the majority of the coefficients, that is, the low spatial frequency coefficients, will be clustered around the DC value. This clustering effect or *energy compaction* is exploited later by the quantization.

The DCT coefficients are quantized so that the non-significant coefficients are set to zero and the remaining ones are represented with a reduced precision. This is achieved by dividing each of the coefficients by an integer taken from a quantizer scale and then rounding the result to the nearest integer. This results in loss of information but also in compression since most of the coefficients in a block are zero.

Finally, the quantized DCT coefficients are further encoded. The non-zero values are encoded using an entropy coding scheme while the zero coefficients are encoded using run-length encoding. Thus, every encoded block is represented with a variable number of bits.

The decoding process follows the inverse steps. First, the variable length codes are decoded to obtain the quantized coefficients. The set of coefficients is then multiplied by the quantizer scale used for encoding (inverse quantization). This operation gives an *approximation* of the original DCT coefficients such that the operation is not exactly an inverse function. Once all the DCT coefficients are available, an inverse DCT is applied to get back to the spatial domain representation of the macroblocks. The inverse DCT operation is given by:

$$f(i,j) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u,v) \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2i+1)v\pi}{2N}\right), \tag{2.3}$$

with $C(x)$ as above and $f(i,j)$, $F(u,v)$ as defined in Eq. 2.1.

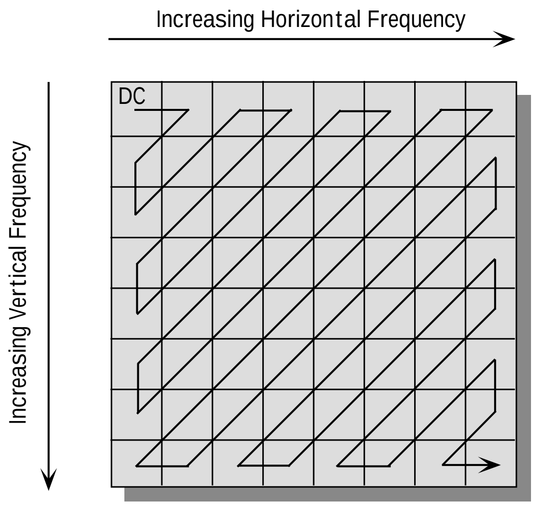


Figure 2.1: *Zigzag scanning pattern in JPEG.*

The video information is represented by a luminance component Y, which contains information about the brightness, and chrominance components for the three basic colors Red Green and Blue (RGB). The four components are combined by substracting the brightness to the color components. The most widely used are the U and V components also known as C_r and C_b to form the YUV color space defined as:

$$\begin{cases} C_r &= w_r(R - Y) \\ C_b &= w_b(B - Y), \end{cases}$$

where w_r and w_b are weighting factors for red and blue difference signals.

2.1.2.2 The JPEG Standard

The Joint Photographic Experts Group defined in 1993 the ISO/IEC 10918 standard format for coding and compression of continuous grayscale and color still pictures. This standard commonly known as JPEG has also been standardized by ITU-T as recommendation T.81 [12].

The JPEG compression algorithm takes advantage of the spatial redundancy of the images. First, a DCT is calculated by scanning the image left to right and top to bottom producing 8×8 blocks of DCT coefficients. In general, the image has a resolution of 8 bits per pixel but higher resolutions are also supported. A quantization is then performed on each of the DCT coefficients. The quantizer scale is in fact a quantization matrix which weights the DCT coefficients. A coarser scale factor is applied with the increasing spatial frequency. This is done to take advantage of the human visual system (HVS) masking property. Indeed, the HVS is less sensitive to degradations in the high spatial frequency range. As a consequence, the lower spatial frequency coefficients are more likely to be nonzero than the higher frequency ones. To further improve compression, the low spatial frequency values are grouped by scanning the DCT coefficients in zigzag order, starting with the DC coefficient and ending with the highest frequency AC coefficients (see Fig. 2.1). This groups the zero values into long runs.

The DC coefficients are encoded separately from the AC ones. DC values tend to be similar between adjacent blocks and so, each DC but the first is encoded differentially. The

AC coefficients are converted into a set of run-length values further encoded using Huffman coding. This results in a set of variable length codes (VLCs).

The output of a JPEG encoder consists therefore in a sequence of variable size intracoded pictures.

The early availability of JPEG hardware and its relative low complexity has been one of the reasons to develop what has been called *Motion JPEG* or MJPEG. It is a very simple extension that considers a video sequence as a set of still images compressed with JPEG. Motion JPEG has never been standardized but is widely available in commercial products. However, some companies have added proprietary control information that makes the flows incompatibles between different systems. Still, MJPEG is used for some high quality video applications, in particular, when lossless compression is required as in telemedicine. The problem with MJPEG is that due to its low compression ratios it requires large amounts of bandwidth for real-time video.

2.1.2.3 The H.261 Standard

The main target of ITU-T recommendation H.261 [11], also known as $p \times 64$, is the video telephony over ISDN lines. H.261 is part of the H.320 group of standards which describes the different components of a video conferencing system and define a narrow-band multimedia terminal.

Unlike JPEG, the H.261 compression algorithm takes advantage of both the spatial and the temporal redundancy of video sequences to achieve high compression ratios. However, the relatively low bit rate available constrains H.261 to support low resolution formats like the Common Intermediate Format (CIF) whose resolution is 352×288 pixels and Quarter CIF (QCIF) whose resolution is 176×144 pixels and cannot deliver video broadcast quality. The maximum frame rate is 30 frames per second but it can be reduced depending on the application and bandwidth availability.

The coding algorithm is based on a DCT transform of the image and a motion-compensated interframe prediction. The data is organized into four 8×8 blocks of luminance one of C_r and one of C_b. *Intracoded* frames, frames which do not have other reference than themselves, are compressed similarly to JPEG. To perform predictions, the intracoded frames are decoded and stored at the encoder for reference.

The macroblocks of the subsequent frames are then *intercoded*. Each macroblock is motion-predicted from the nearby macroblocks in the previous frame. The offset between the previous and current macroblocks is encoded as a motion vector for the macroblock. If the prediction error is less than a certain threshold, no further processing is performed. Otherwise, the error is encoded using the same DCT, quantization and variable length coding.

To help processing at the decoder, the compressed information is encapsulated in a hierarchical way as depicted in Fig. 2.2. The blocks, which are the basic units are made up of DCT coefficients. These variable length blocks are encapsulated into macroblocks which are themselves collected to form a group of blocks (GOB). Finally, a complete picture is made up of several GOBs. This structure requires a syntax to define the containers. The delineation of these containers is done with headers. The headers are used by the decoder to keep the bitstream and the decoder synchronized.

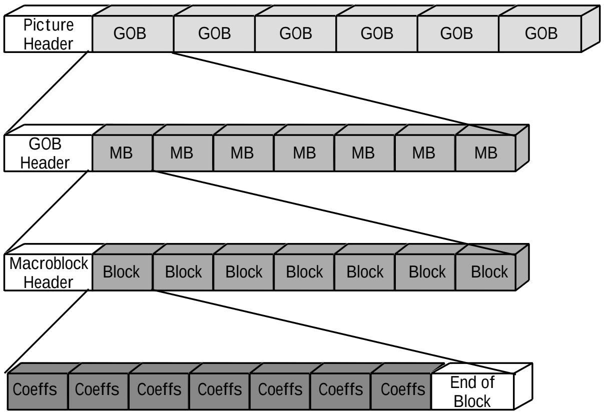


Figure 2.2: *Structure of the H.261 bitstream.*

2.1.2.4 The MPEG-1 Standard

The MPEG-1 standard formally known as ISO/IEC 11172 was originally intended for coding of video and associated audio at bit rates of up to 1.5 Mbit/s. MPEG-1 was originally aimed at digital storage, and in particular compact disks. The main applications targeted by MPEG-1 were prerecorded video, Interactive CD and games. Such applications do not require large image sizes and resolutions. MPEG-1 was optimized to encode noninterlaced video at Source Intermediate Format (SIF) resolutions (352×240) at 30 frames per second, or 352×288 at 25 fps. The bit rate required for these resolutions is about 1.2 Mbit/s. Adding two channels of audio information increased the bit rate to 1.5 Mbit/s. However, MPEG-1 allows to encode much larger picture sizes and correspondingly higher bitrates. The MPEG-1 standard is composed of three main documents specifying the audio encoding, the video encoding and the system layer.

The MPEG-1 video standard [13] specifies the bitstream syntax and the decoding process. In the same way as H.261, MPEG-1 exploits both the spatial and temporal redundancies of motion pictures. The encoding algorithm is similar to H.261. The spatial redundancy is exploited by DCT encoding of 8×8 pixel blocks followed by quantization, zigzag scan and variable length coding. The quantization matrix is also weighted to discard perceptually irrelevant video information. The temporal redundancy is exploited by motion compensation prediction.

The MPEG-1 standard defines three types of encoded pictures: I, P and B pictures.

Intrapictures (I), are intraframe encoded without any temporal reference. I pictures take only advantage of the spatial redundancy and are used as a reference for the other two types of pictures.

Forward predictive pictures or *P pictures* are interframe encoded using motion prediction from the previous reference, either I or P, picture in the sequence. Like in H.261, the luminance component of each macroblock is matched with a 16×16 region in the precedent reference picture. The difference and the motion vectors are then encoded and transmitted. When the motion prediction is not effective, the macroblocks may be intracoded. P pictures are *causal* because they use past references.

Bidirectionally predictive pictures or *B pictures* are interframe encoded using interpolated motion prediction with respect to the immediate precedent reference picture as well as the immediate next reference picture. Each macroblock is compared with its neighboring area in the previous and next I or P picture. The prediction macroblock may be chosen from

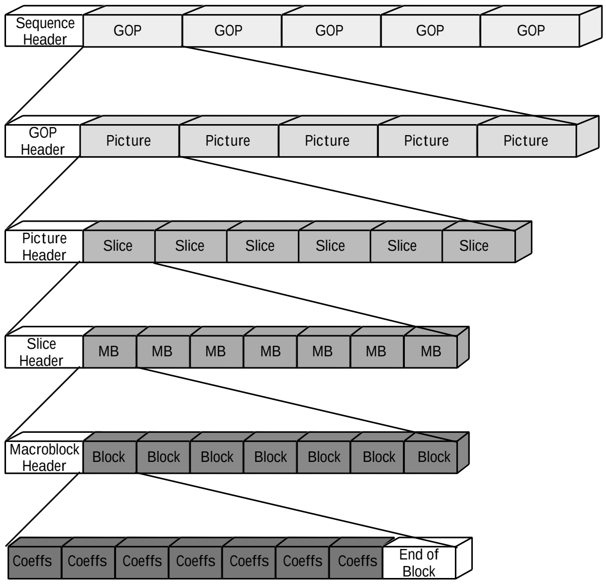


Figure 2.3: *Structure of the MPEG-1 bitstream.*

the forward, the backward or an average prediction. The smallest error prediction is used, or intracoding may also be chosen, if motion prediction is not efficient. The motion vectors, either forward and/or backward, and the prediction error are encoded and transmitted. B pictures are *noncausal* and use two surrounding causally coded pictures for prediction. B pictures are not further used as an encoding reference.

Due to the encoding characteristics, generally B pictures achieve higher compression ratios than P or I pictures at the price of increased encoding complexity.

The three picture classes are grouped together in Group of Pictures (GOP). A GOP defines a structure and ordering of pictures. The first is always an I picture which is followed by a given number of P and B pictures. Figure 2.4 shows a typical GOP example composed of 9 pictures. Two parameters define the GOP: the distance between P pictures and the number of B pictures. GOPs serve as basic access units with the I picture serving as the entry point for random access and to limit the temporal propagation of errors. Like in H.261, the MPEG-1 bitstream is hierarchically organized. The organization is based on a set of headers, the syntactic information. The headers encapsulate further headers but ultimately contain the video or semantic information.

The sequence header encapsulates the whole sequence and contains such information as the picture format. A sequence is made up of pictures organized in GOPs. Each picture is made up of slices which is a macroblock container and finally each macroblock is composed of blocks. The syntactic information is of fixed size and is used by the decoder to resynchronize in case of loss.

The MPEG-1 standard defines a system layer [14] for combining audio, video and user data. The system layer defines a packet structure to multiplex separate data flows into a single bitstream and keeping it synchronized. Each data flow is referred to as an elementary stream. The system layer adds supplementary headers for assisting the decoder in parsing the multiplexed stream as illustrated in Fig. 2.5. To keep the flows synchronized, the system layer generates and inserts *time stamps* into the packets. These time stamps are used for decoding and presentation by the decoder.

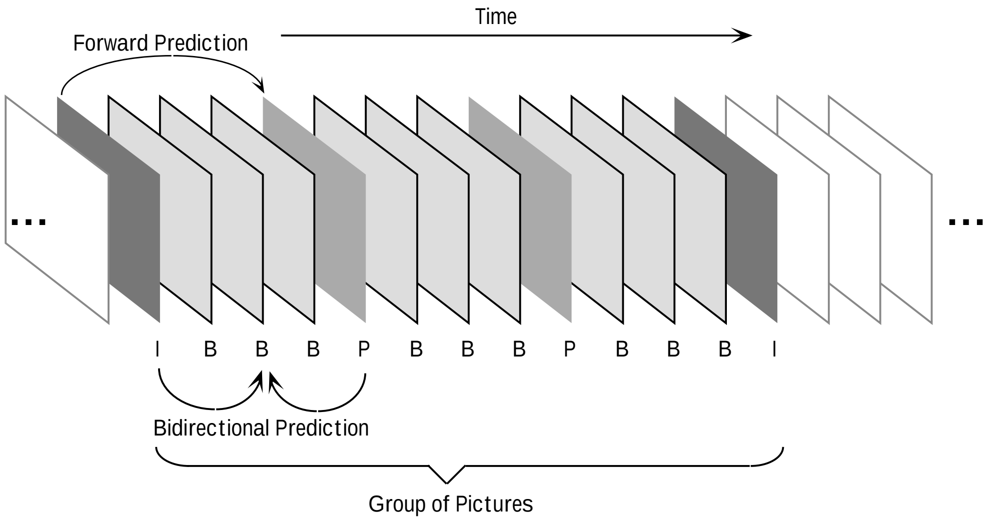


Figure 2.4: *Structure of the MPEG-1 group of pictures.*

Since the MPEG standard was designed as an integrated audiovisual coding and compression algorithm, it also specifies the encoding of audio information. The MPEG-1 audio syntax [15] is based on the perceptual limitations of the human auditory system for encoding and thus much of the compression results from the removal of perceptually irrelevant audio information. MPEG-1 audio allows different sampling rates going up to 48 kHz. It also allows four modes: mono, stereo, dual with separate channels and joint stereo.

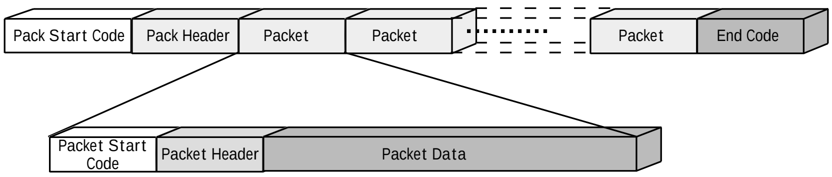


Figure 2.5: *Structure of the MPEG-1 system layer bitstream.*

2.1.2.5 The MPEG-2 Standard

Following MPEG-1, the need arose to compress TV broadcast quality video (ITU-R 601 format). MPEG-1 had some limitations such as the inability to cope with interlaced video and the fact that it was actually optimized to work at around 1.5 Mbit/s. Also, the forthcoming video format for High Definition TV (HDTV) called for a new compression standard capable of achieving high compression ratios on large image high resolution formats.

ISO/IEC developed a second standard officially known as ISO-13818 composed of three main parts: systems [16], video [17] and audio [18]. Parts 1 and 2 are defined as ITU-T standards as well [19, 20].

MPEG-2 extends the functions provided by MPEG-1 to achieve efficient encoding of audiovisual information at a wide range of resolutions and bit rates, but also to provide full interactive multimedia services such as:

- random access for interactive TV

- trick modes are able to provide VCR-like features such as fast forward, reverse play, slow forward play, pause, etc...
- multiple audio and video flows (stereo, multilanguage).

Since MPEG-1 was intended for audiovisual coding for Digital Storage Media (DSM) applications and since DSMs are error-free environments, the MPEG-1 Systems part was not designed to be robust to errors. Also MPEG-1 was not intended for transmission but rather for software processing and thus large variable length packets were used to minimize overhead.

MPEG-2 on the other hand targeted a variety of multimedia applications including transmission. The MPEG-2 Systems [16] was designed to improve error resilience and to carry multiple programs simultaneously without requiring them to have a common time base. To be flexible enough, the MPEG-2 system layer defines two types of streams: The Program Stream and the Transport Stream. The former is analogous to the MPEG-1 system stream with a modified syntax and new functions. It provides compatibility with the MPEG-1 system stream and is intended for DSM and error-free environments. It generates long variable-length packets for software processing with minimal overhead. The latter differs significantly from the MPEG-1 system and the program stream format. The Transport Stream (TS) uses fixed length packets of 188 bytes. It is more suited for hardware processing and error correction. Thus it is well suited for transmission over error-prone channels such as coaxial cable TV networks and packet networks including ATM. The Transport Stream allows for the multiplexing of multiple programs with independent time bases into a single stream.

Both system streams share a common data structure the Packetized Elementary Stream (PES). PES packets are generated by packetizing the continuous stream of compressed data, either audio or video. A Program Stream is generated by simply concatenating PES packets with necessary data to generate a single bitstream. A Transport Stream is obtained by segmenting PES packets into TS packet payloads. A TS packet consists of a 4-byte header followed by 184 bytes of payload as shown in Fig 2.8. In fact, TS payloads could be less than 184 bytes because an adaptation field may be present. Two conditions have to be met to segment PES packets into TS packets:

1. the first byte of each PES packet is always the first byte of a TS packet payload
2. data from only one PES packet should be carried in a TS packet.

Since PES packets are of variable length it is necessary to align PES packets to TS boundaries by padding.

Originally, the MPEG-2 video specification was primarily intended for coding of interlaced video at standard TV resolution in the bit range of 4 to 9 Mbps. However, the scope of MPEG-2 was considerably widened to include higher resolutions and bit rates as well as hierarchical coding. Among the new features included in MPEG-2 are the support of different chrominance sampling modes. A 4:2:2 sampling produces chrominance components with the same vertical and half the horizontal resolution of the luminance components. This provides better color resolution than the standard 4:2:0. A higher resolution 4:4:4 is also supported which provides the same chrominance and luminance resolutions. Another new feature of MPEG-2 is *scalability.* Scalable modes enable video information to be encoded into two or more layers. Four *layered coding* modes are defined in the standard:

- *spatial scalability* allows each frame to be encoded at a range of resolutions that can be built up to the full resolution. This feature could for example be used to transmit video

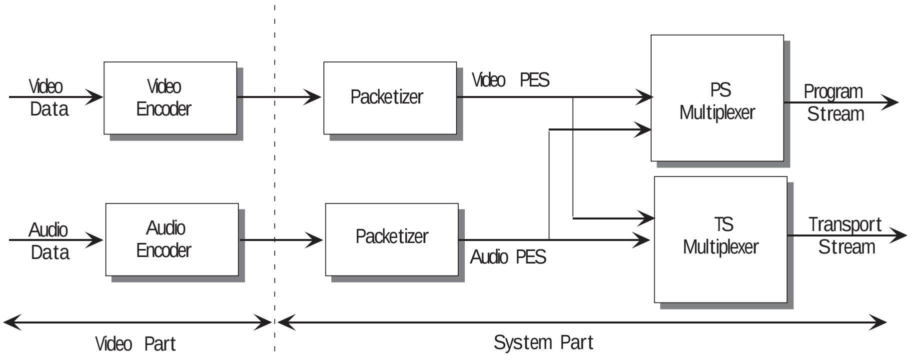


Figure 2.6: *The MPEG-2 reference encoder.*

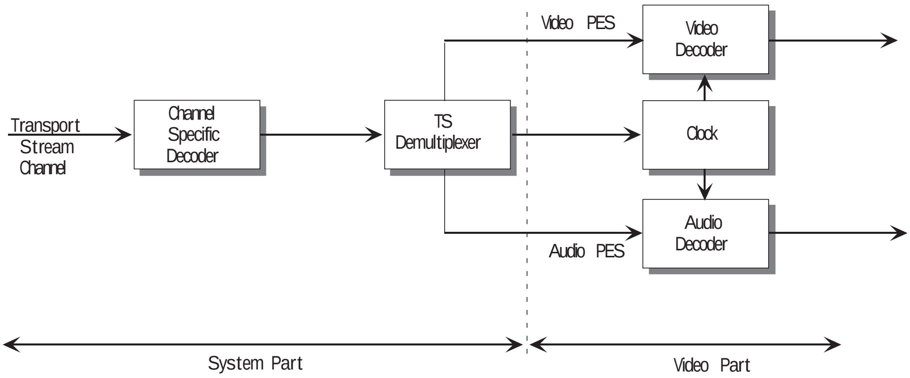


Figure 2.7: *The MPEG-2 reference decoder.*

simultaneously to standard TV terminals as well as HDTV terminals. The base layer could have the standard TV resolution while the enhancement layers could include the information necessary to obtain the full HDTV image

- *data partitioning* enables the coded data to be separated into high and low priority streams. A high priority stream will include the basic information such as motion vectors headers and low frequency DCT coefficients. The low priority ones will contain the remaining information
- *signal to noise ratio scalability (SNR)* allows pictures to be encoded in a basic coarse quality version. The enhancement layers providing information required to decode the full quality image
- *temporal scalability* allows to encode a sequence at different frame rates. The base layer will contain the sequence at a low bit rate and the enhancement layers will contain the remaining frames needed to achieve the full frame rate.

The MPEG-2 video standard does not specify the video encoding. Instead, it specifies the video bitstream syntax and decoding semantics. The basics of MPEG-2 video encoding are the same as for MPEG-1 especially for progressive encoding. It has three types of pictures, I, P and B. Images are DCT transformed, quantized and zigzag scanned. Motion estimation and compensation is then performed. Finally, entropy encoding is done by applying variable length coding to the data. The major differences between both standards reside in the capability that MPEG-2 has to efficiently compress interlaced video. To achieve this, MPEG-2

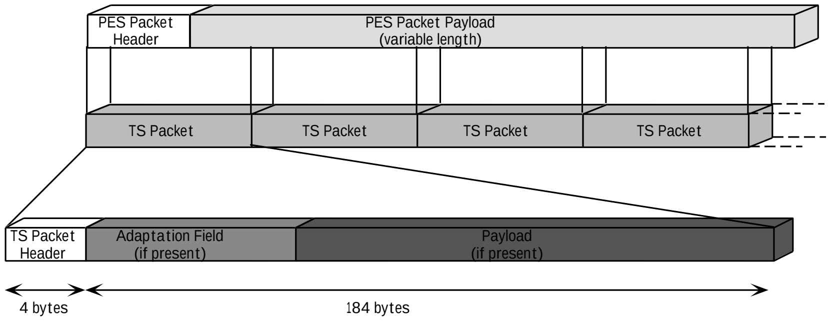


Figure 2.8: *Structure of the MPEG-2 system layer bitstream.*

specifies a choice between two picture structures. *Field-pictures* consist of fields that are coded independently. With *Frame-pictures*, on the other hand, each interlaced field pair is interleaved together into a frame that is then divided into macroblocks and coded. MPEG-2 requires interlaced video to be displayed as alternate top and bottom fields. However, either field can be displayed first within a frame. Another difference due to interlaced video is the *alternate scan* mode offered by MPEG-2. The zigzag reordering (see Fig. 2.1) is basically performed to group spatial frequencies in an increasing order and to generate long runs of zeros. In frame-pictures, adjacent scan lines come from different fields and therefore vertical correlation is reduced when motion in the scene exists. The zigzag scanning may not be optimum in this case. Therefore the encoder may chose in a picture-by-picture basis which mode is better. Another feature to deal with interlaced video is the *field_DCT* coding which reorders the luminance lines to increase energy compaction prior to performing the DCT.

Since MPEG-2 targeted a large range of applications, the defines a set of profiles and levels that provide different encoding parameters.

A profile is a subset of the full MPEG-2 syntax that specifies a particular set of coding features. Each profile is a superset of the preceding profiles. A level specifies a subset of spatial and temporal resolutions which include a large set of image formats. Tables 2.2 and 2.3 summarize the profiles and levels defined in the standard.

Profile	Description
Simple	4:2:0 sampling, I/P pictures only, no scalability
Main	4:2:0 sampling, I P and B pictures, no scalability
SNR	4:2:0 sampling, I P and B pictures, SNR scalability
Spatial	4:2:0 sampling, I P and B pictures, Spatial scalability
High	4:2:0 sampling, I P and B pictures, Temporal scalability

Table 2.2: *MPEG-2 profiles.*

Level	Maximum Resolution
Low	$352 \times 288, 30Hz$
Main	$720 \times 576, 30Hz$
High-1440	$1440 \times 1152, 60Hz$
High	$1920 \times 1152, 60Hz$

Table 2.3: *MPEG-2 levels.*

The standard does not allow all the profile and level combinations, but a subset of them.

In fact, particular profile and level combinations are designed to support particular classes of applications. For example, Main profile and Main level is suited for digital TV applications and is supported by almost all MPEG-2 decoders.

The MPEG-2 standard also includes audio compression. Two modes exist; a backwards compatible (BC) with MPEG-1 audio and a two channel stereo mode equivalent to MPEG-1 audio. The MPEG-2 BC audio mode specifies the encoding of up to six-channel audio sound known as the $\frac{3}{2}$ mode. This mode specifies the encoding of three front and two back channels of audio which are the left, center and right channels plus left and right rear channels. These two channels are used for surround audio. The sixth channel which is optional is the low-frequency enhancement channel. Actually the MPEG-2 standard allows several combinations of multichannel audio to be encoded. In many cases, a dual mode is used. This consists of using two pairs of stereo channels for e.g bilingual programs. However, the BC mode does not achieve the best audio quality possible. A non-compatible mode able to achieve state-of-the-art five channel audio quality is under consideration among which the AC-3 system from Dolby.

2.1.2.6 H.263

The development of modems allowing transmission in the range of 28 to 33 kbps over PSTN paved the way for the development of an improved version of H.261 for conversational video services at very low bit rates. ITU-T released in 1996 the H.263 standard [21]. H.263 is an enhanced version of H.261 which achieves higher coding efficiency. Thus, like H.261, H.263 is a video coding standard and does not specify audio or systems multiplex which can be chosen from related ITU-T standards.

The major improvements over H.261 consist of introducing a half-pixel motion compensation (like in MPEG-1 and 2) which allows more precise prediction. Another difference appears in the utilization of GOB headers which is optional in the new standard thus reducing overhead. To allow further improved performance, the standard provides four negotiable options which can be used together or separately. Among these options, the utilization of syntax-based arithmetic coding instead of Huffman coding is available. The coder can also generate B pictures, which are unavailable in H.261, but unlike MPEG standards, a B picture is coded together with a P picture as a single PB-picture unit.

H.263 has no bitrate limitation and is able to support five picture formats from sub-QCIF to 16CIF which actually gives a picture format of up to 1408×1152 luminance resolution.

2.1.2.7 MPEG-4

Even if MPEG-2 is a very generic compression algorithm, multimedia applications become more and more complex. The introduction of virtual reality requires efficient coding of synthetic models. Also, foreseen applications may contain multiviewpoint scenes and graphics.

The issue is that new applications as well as the development of wireless communications show that MPEG-2 is not generic enough to efficiently handle the forthcoming new applications and services.

The aim of the MPEG-4 working group is to produce an even more generic standard which would be efficient, flexible and extendable in the future. Albeit, the original target of the MPEG-4 committee was the development of a coding standard for very low bit rate applications, the focus changed. Indeed, the MPEG-4 working group desire was to incorporate fundamental advances in video coding technology to achieve very low bit rates such as wavelet or fractal coding. The current state of that research is not mature enough to be incorporated into a standard.

The basic functionality classes that the MPEG-4 standard will provide are [22]:

- content-based coding and manipulation of multimedia data. This allows to interact with objects in a scene. It also embraces all aspects of data access via the Web browser paradigm
- improved coding efficiency for storage and transmission over heterogeneous networks as well as the coding of multiple concurrent streams for multiviewpoint scenes
- error robustness owing to an increasing focus on mobile communications
- content-based scalability.

To achieve such capabilities, MPEG-4 has to provide the possibility to access not only pictures but also regions or objects within a picture. This lead to the concept of Video Object Planes (VOPs). A VOP can be a semantic object that is represented by texture variations and shape information. Each VOP may contain a single object within a scene. For example, in a videoconference sequence, the head and shoulders view of the remote speaker will be contained in a VOP and the background in a second one.

The MPEG-4 working group has based the initial support of low bit rate video on the H.263 algorithm. The MPEG-4 standard will not define a coding algorithm but rather will define a set of tools from which the applications will be able to select and download as required. This structure will be flexible enough to be expandable because it allows any new future coding technique to be incorporated as a new tool into the generic codec.

The advantage of this architecture is that it allows each VOP in a sequence to be coded separately and with a different algorithm. Each algorithm could be chosen to achieve the best compression possible according to the VOP type. This architecture will be supported by the MPEG-4 Syntax Description Language [23]. The MSDL could be considered as a system layer addressing the system capabilities needed to support the MPEG-4 functionality classes. MSDL defines three types of decoder programmability to support flexibility and extensibility which are:

- level 0 (nonprogrammable) decoder incorporates a prespecified set of standardized algorithms
- level 1 (flexible) decoder incorporates a prespecified set of standardized tools which can be flexibly configured into an algorithm at setup phase
- level 2 (extensible) decoder provides a mechanism for the encoder to download new tools and algorithms.

Currently, the MPEG-4 standard is in the verification test phase. The final document is expected to be approved by ISO in 1998.

2.1.2.8 Remarks

The drawback with compression algorithms is that the less the redundancy the more sensitive to loss the data streams become. This adds an additional constraint to the network. Users are tolerant to image errors, however, errors in compressed streams quickly become noticeable and therefore annoying to the user.

Another issue related to compression algorithms is the nature of the data stream delivered to the network. Compressed video streams can be categorized into two classes: *constant quality* and *variable quality* streams. The former, matches better the nature of video information which is basically variable. An increase in motion or a scene change will fundamentally modify the amount of information to transmit. Several studies have been done to characterize video sources. Recent contributions [24, 25] have shown that VBR video has long range dependencies and a self-similar nature. The latter, which today is the more widely used delivers a Constant Bit Rate (CBR) stream which is much more easy to transmit and which presents several advantages for the application. Timing and delay aspects are simplified. However, to achieve a constant data rate, the quality has to be modified. The way this is performed is by adding a rate control to the encoder. A rate control based on the monitoring of an output buffer adjusts the quantization factor to achieve a constant output. From a perceptual point of view, scene changes or high motion sequences may show degradations such as blocking effects and mosquito noise. Another inconvenient of CBR streams is that they do not fully exploit the capacity of packet networks.

The area of audiovisual compression is still developing and new algorithms are under consideration for the MPEG-4 standard. The trend is to further increase the compression ratios. This has two opposed outcomes: firstly, if high quality video could be achieved with very low bit rates then the probabilities of having losses or errors in networks becomes very small. Secondly, by further increasing compression, more and more redundancy is removed which increases the impact that losses could have onto video.

Video compression algorithms are one of the most important enablers of multimedia communications. By removing redundancy the applications require relatively little bandwidth but simultaneously become much more sensitive to loss in addition to the sensitivity to delay. Therefore the transmission has to be more reliable which adds a supplementary constraint. This requires from the network to deliver multimedia data in a *timely* and *reliable* manner in order to achieve from the user's perspective a good QoS.

2.2 Network Functions

This convergence of multiple services and traffic profiles requires a generic network capable of handling multiple classes of service. Real-time multimedia applications need to be handled by the network in a different way than traditional data applications.

The concept able to cover such a wide spectrum of requirements is *fast packet switching* (FPS). Basically it is a packet network with minimal network layer functionality which allows for high data transfer rates. ATM as a FPS technology was chosen as the core infrastructure for the future B-ISDN because it provides a flexible and efficient delivery of any type of data accommodating many different bandwidth, delay and traffic characteristics. LAN interconnection, distributed interactive multimedia services including conversational and browsing and even broadcast services are some of the applications that will be used by corporate as well as residential users. It is a broadband, low delay technology which can be deployed in public as well as in private networks.

The next sections briefly describe the ATM technology principles, and what the current offerings from the network point of view are related to multimedia.

2.2.1 Basic ATM Concepts

ATM is a *packet network* and therefore it does not use dedicated circuits for the communications. However, it is well known that packet networks can hardly achieve good QoS in particular in terms of delay. To overcome this problem, ATM works in a *connection-oriented* mode. A call setup is performed prior to the establishment of the connection. Resources, if available, are allocated and a path is established. This guarantees cell sequencing at the receiver and the same delay for each cell (with small variations). The allocation is done based on the *traffic contract* which specifies the source parameters and a set of QoS parameters requested by the user. The advantage of this technique is an efficient utilization of the network resources because they are shared.

To allow low delays, as required by real-time applications, ATM uses small packets. This reduces the queueing delays because buffers do not need to be large. In addition to being small the ATM cells are of fixed size which helps to reduce the delay jitter. To allow fast switching not only the packet size has an influence. Also, the processing delays are very important. To reduce them, the information carried by the ATM cell headers has been reduced to a minimum which includes almost only information for routing.

The third element to achieve low delays is the avoidance of any flow control or error correction mechanism. No flow control is applied in a link so overload may occur and cells may be lost. In other packet networks error correction is performed by retransmission (e.g. ARQ). This leads to large delays whose lower bound is the round trip time which is too large for interactive applications. ATM does not take any corrective actions. It is up to the higher layers to handle errors.

One of the problems of ATM as described above is that it assumes that every sender is able to specify, at least to some extent, the behavior of the source. This was early identified as a limitation in particular for data transfer sources. Traffic such as the one generated by a LAN-to-LAN interconnection is almost completely unpredictable and extremely bursty in nature [26]. On the other side, this kind of traffic does not have any particular delay requirement. Also, the high variability of this traffic makes difficult to reserve resources without any waste of bandwidth.

The Available Bit Rate service (ABR) described in Sec. 2.2.4 was specified to overcome this problem. The philosophy of ABR is closer to legacy data networks than to the original concept of ATM. ABR provides rapid access to unused network bandwidth whenever it is available. ABR connections require maximum and minimum cell rate values. The main principle of ABR is to provide a rate-based closed loop per-connection flow control which uses feedback information to regulate the source rate between these two values. As a consequence, no loss is to be expected as far as the source keeps its traffic conforming to the contract. This concept changed several things in the architecture of ATM network elements. Since a closed loop flow control is used, the network has to be able to store a relatively large number of cells. Large buffers are therefore needed in the network that lead to large delays. This is not a problem for the applications foreseen to use ABR but goes against the principle of low delay mentioned earlier. To be able to provide both services it becomes necessary to have separate buffers; small buffers for delay sensitive applications and large buffers for loss sensitive applications which will use ABR.

The following sections are not intended to be an exhaustive description of ATM but a short introduction to the basic concepts which are related to this thesis. Several books and papers are devoted to the task of introducing and explaining ATM in detail [27, 28]. Several standard bodies like the European Telecommunications Standardization Institute (ETSI) and the ATM Forum publish ATM specifications. Of course the complete specification of ATM can be found in the ultimate reference, the I series of recommendations, developed and published by the International Telecommunications Union (ITU-T).

2.2.2 ATM and the OSI Reference Model

The ATM logical protocol model is composed of three planes like the Narrowband ISDN (NISDN) [1]. The user plane is responsible for the transfer of user information, the control plane supports the signaling for call control and connection functions and the management plane provides network supervision functions. The latter is further subdivided into a layer management and a plane management. As far as ATM is concerned, the user plane is composed of three layers; the physical layer, the ATM layer and the ATM Adaptation Layer (AAL) (Fig. 2.9). Even though the OSI model is still the reference for layered protocols, there is no exact mapping between the ATM protocol layers and the OSI model. It can however be said that ATM covers layers one and two, the physical and the data link layer, but also covers functions of layer three, the network layer. The ATM layer performs such tasks as segmentation and reassembly, which may be considered as the equivalent framing function performed by the data link layer in the OSI model, However, what makes the ATM layer behave also like a network layer is its hierarchical address space and its routing functions.

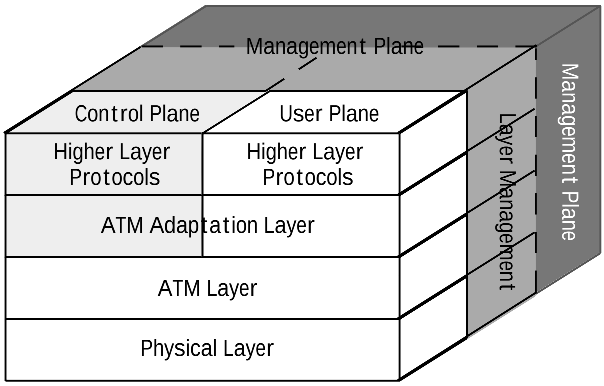


Figure 2.9: *B-ISDN ATM protocol reference model.*

2.2.2.1 The Physical Layer

The physical layer's [29] main task is the transport of data over a specific physical medium (e.g, optical or electrical). ATM was designed to be deployed over several physical media. To get rid of the specifics of a single medium, the functions have been splitted into two sublayers, namely the Physical Medium Dependent (PMD) sublayer and the Transmission Convergence (TC) sublayer. The PMD supports the medium specific functions (like modulation). The TC sublayer adapts the information to a given transmission system such as the Synchronous Optical Network (SONET), Synchronous Digital Hierarchy (SDH) [30] or Plesiochronous Digital Hierarchy (PDH). It is also in charge of the generation of the Header Error Control (HEC) parity check data used to protect the cell header and perform the cell delineation function. Also, since the user transmits in an asynchronous way, this layer has to adapt the user rate to the available transmission rate by generating *idle* cells to keep the line in a synchronous state.

2.2.2.2 The ATM Layer

The two sublayer structure of the physical layer hides to the upper layers the characteristics of the physical medium used for transmission. The ATM layer is therefore independent of the underlying medium and does not perform any medium specific functions.

The ATM layer [31] receives data from the upper layer as well as from the physical layer which may belong to different connections identified by their Virtual Path Identifier and Virtual Channel Identifier (VPI/VCI) values. Hence, one of the functions of this layer is to multiplex the data cells coming from different connections for transmission over a single physical line and conversely demultiplex the data coming from the physical layer for distribution to the different connections. It also may modify the VPI/VCI fields in switches or cross-connects stages. The other main function of the ATM layer is the generation of the ATM cell header. Figure 2.10 depicts the different fields of the header at the User-to-Network Interface (UNI). The 5 octets contain 6 fields generated by this layer. The VPI and VCI fields identify the connection and are used for switching. The Payload Type (PT) field intended to be used for a generic AAL-PDU packet delineation is currently used only by ATM Adaptation Layer 5. The Cell Loss Priority (CLP) bit specifies whether a cell should be discarded preferably in case of congestion or not which is generally referred to as high and low priority cells [32]. Finally, the HEC field is created but reset because as explained in Sec. 2.2.2.1 it is filled by the physical layer.

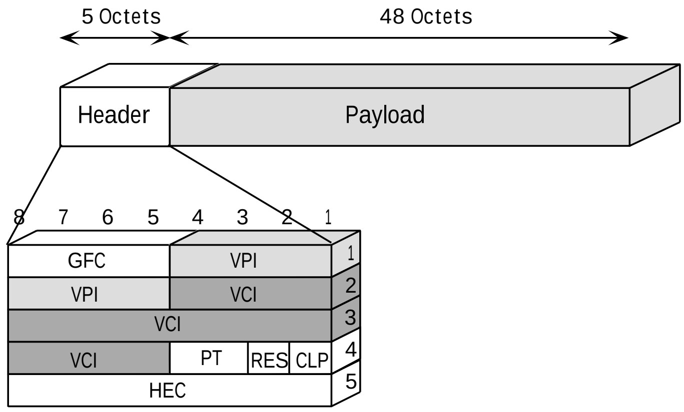


Figure 2.10: *ATM cell structure.*

2.2.2.3 The ATM Adaptation Layer

The AAL enhances the services provided by the ATM layer to support the requirements of a specific service. Each AAL provides specific functions tailored to a set of classes of service. To fit to the classes of service, 5 AALs have been defined albeit not all have been specified up to date [33, 34]. This service classification is based on three attributes, namely; timing relation between source and destination, bit rate and connection mode (see Tbl 2.4). Among all the combinations four classes of service have been defined:

- Class A: is a connection-oriented CBR ATM transport service with an end-to-end timing relation. This service typically covers circuit emulation and real-time CBR multimedia applications. AAL1 provides the required services for these applications.

- Class B: is a connection-oriented VBR service with an end-to-end timing relation. The applications targeted by this class of service are real-time VBR video multimedia applications. AAL2 should provide the required functions for such applications. However by the time of writing ITU-T Study Group 13 has not developed any specification.

- Class C: is a connection-oriented VBR ATM transport service with no specific end-to-end timing requirements. This service is intended for connection-oriented data transmission such as Frame Relay. AAL3/4 and AAL5 provide the required functions for such service.

- Class D: is a connectionless VBR service with no specific end-to-end timing requirements. This service is intended for connectionless data transmission such as SMDS. LAN interconnection is a typical application for this class of service. AAL3/4 provides the required functionalities for such service.

Note that a VBR connection does not necessarily mean that the source generates a variable bit rate. In fact, VBR traffic encompasses CBR traffic. Any CBR application could use a VBR connection. The particularity of AAL1 concerns the timing requirements which are not necessarily covered by other classes but class B. Other reasons such as the scheduling are a factor that call for the utilization of AAL1 when true CBR is required.

A fifth class of service called class X is an unrestricted service in which the user specifies only bandwidth and QoS parameters (i.e. cell-relay service). AAL5 supports this class of service.

	Class A	Class B	Class C	Class D	Class X
	AAL1	AAL2	AAL3/4 AAL5	AAL3/4	user defined
App Type	Voice, CES	Video	Data (e.g FR)	Data (e.g. SMDS)	user defined
Timing	Required	Required	Not Required	Not Required	user defined
Bit Rate	Constant	Variable	Variable	Variable	CBR & VBR
Connection Mode	Conn-Oriented	Conn-Oriented	Conn-Oriented	Connectionless	Conn-Oriented

Table 2.4: *Service classes for the AAL.*

Recent work of Study Group 13 Question 6 (SG-13 Q6) [35, 36] has decided to abandon this classification estimating that it is not suitable anymore. It was originally intended to be used as a guide for AAL development. However, further developments have shown the obsolescence of this classification. In particular, the development of a new AAL provisionally named AAL-CU, which stands for Composite User Information, that provides the multiplexed transport of variable length (short) information with variable bit rate under tight delay constraints is not covered by any of the service classes defined in I.362. Therefore, SG-13 Q.6 has decided to revise I.362 and to abandon this classification without replacing it [36]. In addition, the mapping between service classes and AAL seems not to be adequate anymore in the sense that e.g. AAL5 is now also used for the transport of class A services.

In the same way as the physical layer, the AAL is further subdivided into two sublayers. The *Segmentation and Reassembly* (SAR) sublayer segments the AAL-PDUs into 48 octets cells so called SAR-PDUs. These cells are then sent to the ATM layer which prepends the ATM header. It also performs the opposite function consisting of reassembling the incoming cells into AAL-PDUs. The *Convergence Sublayer* (CS) provides an AAL *Service Access Point* (AAL-SAP) to the layer above and is service dependent. The SAR and the CS sublayers may in some cases be empty.

The next section describes in detail the specific functions of each AAL. We discuss the suitability of each of the layers for the transport of real-time multimedia data.

2.2.3 ATM Adaptation Layers

2.2.3.1 AAL type 1

The services provided by AAL type 1 to the user layer are:

- the transfer of Service Data Units (SDU) with a constant source bit rate and the delivery of them with the same bit rate
- the transfer of timing information between source and destination
- the transfer of structure information between source and destination
- the indication of lost or errored information which is not recovered by the AAL.

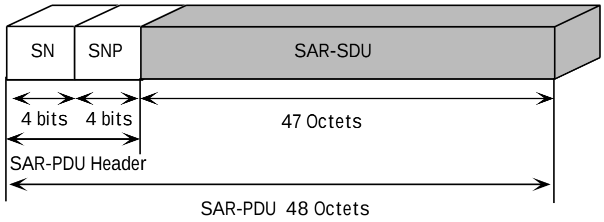


Figure 2.11: *AAL type 1 SAR-PDU format.*

Some of these services are not mandatory and will therefore depend on the CS. The SAR sublayer basically provides a single function which is the mapping between CS-PDUs into SAR-PDUs whose structure is depicted in Fig. 2.11. 47 octets SAR-SDUs are accepted by the SAR sublayer which prepends or removes an octet containing a sequence number (SN) and a protection field (SNP). The 48 octets SAR-PDUs are passed to the ATM layer. The SAR sublayer has also the capability to indicate the existence of a CS function to its peer entity. The convergence sublayer provides a set of functions necessary to cover the aforementioned services which are:

- handling of Cell Delay Variation (CDV). This is done by buffering the CS-PDUs at the receiver

- handling of lost and misinserted cells. Misinserted cells are considered as lost and are dropped. When cells are lost, to maintain the continuity, dummy cells are inserted. Also a forward error correction mechanism combined with an octet interleaver may be applied for further cell loss correction especially for the transmission of video and high quality audio signals

- source clock frequency recovery at the receiver. When no common network clock reference is available, clock information is sent via the Synchronous Residual Time Stamp (SRTS) method

- recovery of source data structure. AAL1 supports a Structured Data Transfer (SDT) which allows for the transmission of octet structured information such as 64 kbps ISDN data

- handling of cell payload assembly delay.

Among these functions the handling of CDV, handling and correction of lost and misinserted cells and the handling of the timing relation are considered as mandatory to support the transport of video signals for interactive and distributive services. However, the FEC mechanism described in the recommendation is specified for *unidirectional services* only.

To improve the suitability of AAL1 for the transport of real-time multimedia applications, ITU-T SG-13 Q.6 [37], introduced modifications into the AAL1 functions. In particular, delay problems regarding the utilization of the octet interleaver were identified. To reduce the delay incurred by the accumulation of data cells in the matrix interleaver, a *short interleaver* was specified as being the correction method for bit errors and cell losses with *delay restrictions* (e.g. audiovisual applications). The problem with this new method is that the overhead is increased from 3.1% to 6.3%.

2.2.3.2 AAL type 2

AAL type 2 was foreseen as the layer for VBR real-time services. However, ITU-T has not yet reached an agreement on this AAL specification. The definition of the services that AAL2 should provide are:

- transfer of SDUs with a variable source bit rate
- transfer of timing information between source and destination
- indication of lost or errored information which is not recovered by the AAL

In the last draft version of recommendation I.363.1 [37], the following functions are to be provided to support AAL2 services:

- segmentation and reassembly of user information
- handling of Cell Delay Variation (CDV)
- handling of lost and misinserted cells
- source clock frequency recovery at the receiver.

These services are close to those offered by AAL1 excepted the SDT service. However, in the current specification there is no description of the cell structure and both the SAR protocol and the CS section are labeled for further study. Actually, it seems that no real specification of the AAL2 will be done by ITU-T. Recent developments, in particular the ATM Forum specification for video transmission over ATM, make use of AAL5 for the transport of real-time VBR (rt-VBR) albeit only the first item in the services to be provided by the AAL are covered since as it will be described later, no timing information is transferred and no handling of cell loss is done by AAL5.

2.2.3.3 AAL type 3/4

Originally AAL type 3 was developed for *connection-oriented* data transfer and AAL type 4 for *connectionless* data transfer. Both specifications were enhanced and merged due to their similarities. AAL3/4 as it was renamed provides two modes of service which are the *streaming mode* and the *message mode*. In streaming mode, an AAL-SDU may be transferred in one or more AAL Interface Data Units (AAL-IDUs). An AAL-IDU is the data unit passed between AAL sublayers which are described later. This mode also allows for transferring an SDU before it is completely available at the sender, a feature that has been used for the transfer of connectionless data via connectionless servers [38, 39]. In message mode the AAL-SDU is transmitted as a single AAL-IDU. The message mode allows for the transfer of fixed or variable length packets while the streaming mode allows for the transfer of variable length packets only.

Both modes of service may offer one of the following operational procedures:

- assured operation: every SDU is delivered with exactly the same data content as the user sent. Any corrupted or lost CS-PDU is retransmitted. Flow control is mandatory
- non-Assured operation: lost or corrupted AAL-SDUs are not corrected by retransmission. An optional error delivery may be used to deliver corrupted AAL-SDUs to the user. Flow control is optional.

All these possibilities of operation of AAL3/4 entail an extra overhead at the SAR sublayer as depicted in Fig. 2.12 leading to 44-octet payloads. The functions of this sublayer are:

- segmentation and reassembly of variable length SAR-SDUs into 44-octet SAR-PDUs (see Fig. 2.12). Multiple SAR-SDUs can be transmitted concurrently over a single ATM layer connection

- preservation of SAR-SDU: two fields in the SAR-PDU support this feature. The Segment Type (ST) indicates if it is the first, a middle or the last segment of the message. Also a single segment message can be indicated. The Length Indicator (LI) is used to align the last SAR-PDU to the 44-octet payloads

- error detection and handling: detection of bit errors and cell losses is possible via a 10-bit CRC and a sequence number

- SAR-SDU sequence integrity: assures that the SAR-SDUs are delivered in sequence within one SAR connection

- multiplexing/demultiplexing: provides the multiplexing and demultiplexing of multiple SAR connections over a single ATM layer connection (identified by a single VPI/VCI). This feature is supported by the *Multiplexing Identifier* (MID) field in the SAR-PDU which allows for 2^{10} AAL user-to-user connections over a single ATM layer connection for connection-oriented data communications. Connectionless communications use the MID to interleave SAR-PDUs belonging to different CPCS-PDUs. This feature is used to interconnect LANs over ATM as illustrated in Fig. 2.13

- abort: provides for aborting a partially transmitted SAR-SDU. When cell losses are detected the transmission of the rest of the SDU may be aborted to avoid inefficient use of the bandwidth.

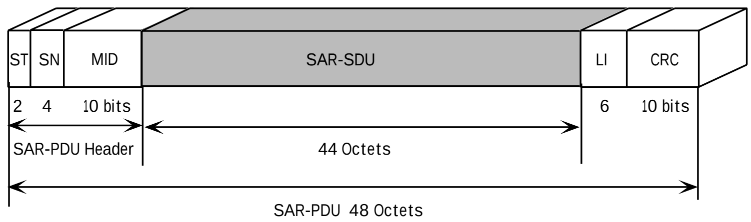


Figure 2.12: *AAL3/4 SAR-PDU format.*

The joint AAL3/4 provides a basic set of functionalities to support a connection-oriented frame relaying telecommunication service in class C as well as a connectionless network access protocol (CLNAP) service in class D. The reason for this common AAL was the identification of identical Convergence Sublayer functions. Since, originally, each layer provided services for different classes, the CS was further subdivided into a *Common Part Convergence sublayer* (CPCS) and a *Service Specific Convergence Sublayer* (SSCS). The CPCS provides all the common functions while the SSCSs are specific to a service class. Class D services do not need any further function and therefore no SSCS is applied. The definition of an SSCS for class C is still under study today. The services provided by the CPCS are:

- non-assured transmission of variable size user data frames (up to 65535 octets)

- multiple CPCS connections may be established between peer entities

- error detection and indication (cell loss or gain)

- CPCS-SDU sequence integrity.

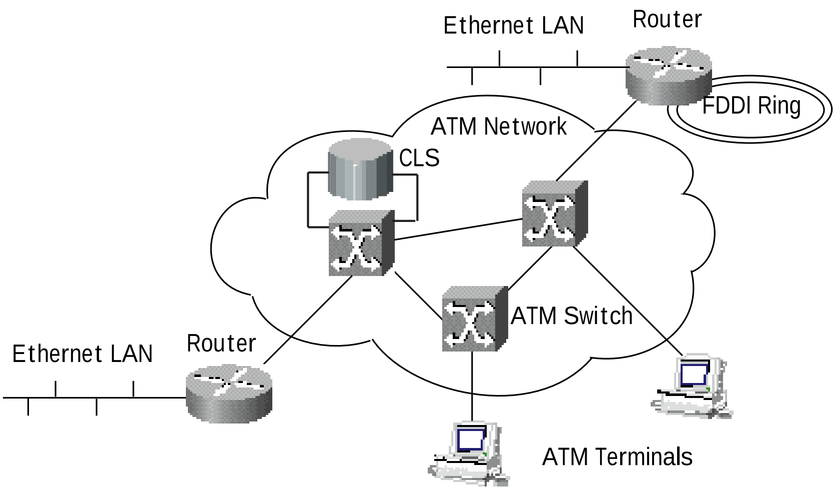


Figure 2.13: *LAN to LAN interconnection over ATM.*

The functions provided by the CPCS to support these services are:

- preservation of CPCS-SDU: provides the CPCS-SDU delineation
- error detection and handling: discards the corrupted CPCS-SDUs or optionally delivers the corrupted packets to the SSCS
- buffer allocation: provides the maximum buffering requirements to receive CPCS-PDUs to the receiving peer entity
- abort: provides the means to abort partially transmitted CPCS-SDUs.

As for the SAR-PDUs, all the features supported by the CPCS require some overhead. Four octets are added in the header and four more in the trailer (Fig. 2.14). Within the header, the Btag field is used in association with the Etag field. They are used as delineation patterns for the CPCS-PDUs. Both fields have the same value within a CPCS-PDU, values that are changed for successive CPCS-PDU. The Buffer Allocation Size field (BASize) is used by the receiver to allocate the maximum buffer requirement to receive a CPCS-PDU. In message mode, the value is equal to the CPCS-PDU payload length while in streaming mode, it is set equal to or greater than the payload size. The Padding (PAD) field is used to align the CPCS-PDUs to the SAR-SDU boundaries in association with the Alignment (AL) field. Finally, the Length Indicator (LI) field is used to encode the size of the CPCS-PDU, information that is also used for error detection.

AAL3/4 is not suited for the transport of class B services and as a matter of fact it is widely replaced by AAL5 for class C services also. The advantage of AAL3/4 is its capability of handling multipoint-to-multipoint connections due to the multiplexing provided by the MID field, feature not available in other AALs. The evident drawback is its large overhead and complexity.

2.2.3.4 AAL type 5

AAL5 was specified by the ATM Forum to overcome the complexity and high overhead introduced by AAL3/4 for data communications. The objective of AAL5 is to offer a service with reduced overhead and a better error detection below the CPCS. Besides, AAL5 CPCS

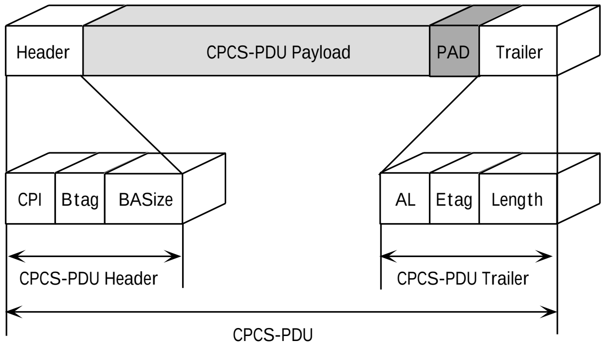


Figure 2.14: *AAL3/4 CPCS-PDU format.*

offers the same services than AAL3/4 but the multiplexing capability. If multiplexing is required, it will be performed at the SSCS.

The SAR sublayer functions of AAL5 are reduced compared to those of AAL3/4. It accepts SAR-SDUs whose size is a multiple of 48 octets. The error detection is not done partly at the SAR sublayer but only at the CPCS. Moreover, no multiplexing service is provided by the CPCS. Therefore, there is no need for a sequence number, a CRC, a MID or a Length Indicator leading to a SAR-PDU payload of 48 octets. The SAR-SDU delineation information is also conveyed by the PTI field of the ATM cell header. Thus, no overhead is added at the SAR sublayer which provides the following functions:

- preservation of SAR-SDU: delineates the SAR-SDUs by inserting an end of SDU indication in the ATM user-to-user information field encoded in the PTI field of the ATM cell header
- handling of congestion information: passes congestion information between the layers above and below the SAR sublayer
- handling of loss priority information: passes cell loss priority information between the layers above and below the SAR sublayer.

Since the convergence sublayer structure of AAL5 is picked up from AAL3/4, we find again the two sublayer structure. The CPCS services are the same as AAL3/4 without multiplexing capability. Therefore, the functions provided are in part different leading to a different CPCS-PDU format (Fig. 2.15). The services to be provided by AAL5 are:

- non-assured transmission of variable size user data frames (up to 65535 bytes)
- the CPCS connection will be established between peer entities by the management or the control plane
- error detection and indication (cell loss or gain)
- CPCS-SDU sequence integrity on each CPCS connection.

The functions needed by the CPCS sublayer to support these services are:

- preservation of CPCS-SDU: provides the CPCS-SDU delineation
- preservation of CPCS user-to-user information: provides the transparent transfer of the user-to-user information included in the CPCS-PDU trailer

- error detection and handling: discards the corrupted CPCS-SDUs or optionally delivers the corrupted packets to the SSCS
- abort: provides the means to abort partially transmitted CPCS-SDUs. The function is indicated in the Length field
- padding: the SAR accepts only multiples of 48 octets so the CPCS has to do the alignment
- handling of congestion information: passes congestion information between the layers above and below the CPCS sublayer
- handling of loss priority information: passes cell loss priority information between the layers above and below the CPCS sublayer.

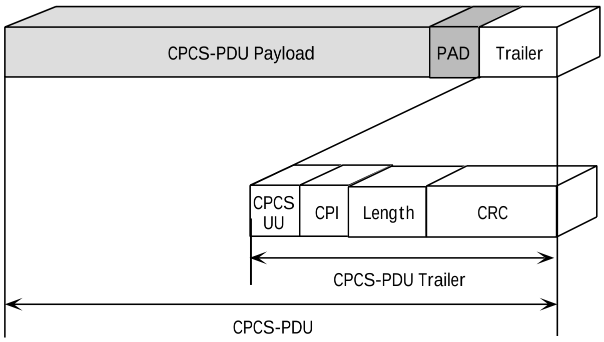


Figure 2.15: *AAL5 CPCS-PDU format.*

Since the main functions of this layer are in the CPCS the overhead is found in the CPCS-PDU (Fig. 2.15). The 8 octet trailer contains four fields: The CPCS User-to-User (CPCS-UU) indication contains user information to be carried end-to-end. The Common Part Indicator (CPI) field is used for alignment purposes. The Length Indicator (LI) contains the size of the CPCS-SDU. This information is used by the receiver to check for cell loss or gain. Finally the Cyclic Redundancy Code (CRC) field contains a 32 bit parity check used to monitor bit errors. The CRC code protects the CPCS-SDU plus the other three fields of the trailer.

In July 1995, SG-13 Q.6 finalized a set of protocols and procedures to provide an option of corrupted data delivery to AAL5. This option offers in the non-assured mode of operation the transfer of corrupted CPCS-SDUs to the layer above. However, in general this option cannot easily be used since the information transmitted with the corrupted SDU is very incomplete. Six cases have been defined as detectable errors. Among these, the case when an incomplete SDU has been detected due to cell losses within the PDU is not considered. Basically, the cases that allow for passing corrupted PDUs are two: CRC reminder error, which is due to bit errors or cell misinsertion, concatenation of PDUs due to End of Message (EOM) cells lost and reassembly timer expiry.

Even though improvements have been done to AAL5 regarding the delivery of corrupted data, it is clear that it is not suited for class B services. It does neither fulfill the timeliness nor the reliability criteria derived in section 2.1.2. However, it has the advantage of handling point-to-multipoint connections which is an advantage for video broadcast applications. A second and even more important advantage of AAL5 is its ubiquity. Since AAL5 is used to transfer signaling information it is implemented in all ATM boards. Therefore, the deployment of new services and applications that make use of AAL5 is straightforward.

2.2.3.5 AAL Composite User information

Recently, ITU has identified the need for an AAL which should provide functions to support mobile communications. The functions to be provided by the AAL are:

- low cell packing/unpacking delay by supporting short length of packets (e.g. up to 64 octets) of variable length, assigned for each user information stream

- efficient use of ATM cell payload, compared to the partially filled cell method with one user information occupying portion of the cell payload and dummy octets occupying the rest of the cell payload.

These features are not available in the existing AALs described earlier. The AAL-CU[1] which stands for Composite User information makes it possible to use an ATM connection efficiently. Good efficiency is achieved when the same connection, simultaneously, can carry a significant number of short packets with low bit rates. The kind of applications able to generate such type of traffic are mobile communications with speech compression. However, this AAL is not aimed at carrying traffic between end-users, since ATM is not terminated at the mobile terminal but for trunking between base stations and mobile switching centers.

It is worth to note that the development of new AALs is still carried on by ITU if the requirements of new major applications are not fulfilled by the available set of AAL services and functions.

2.2.4 Traffic Enforcement and QoS

One of the main advantages of ATM is its capability to guarantee the appropriate QoS to a wide variety of services and applications. Achieving certain network performance objectives [40] depends on the network's ability to perform two main tasks; proper resource allocation and control of the incoming traffic. To correctly allocate resources, the network has to know the incoming traffic characteristics and resource availability. ATM provides functions to meet network performance objectives based on both the traffic characteristics and the availability of resources. These functions are described in recommendation I.371 [41] as the *Traffic Control and Congestion Control functions for B-ISDN*. Their primary role is to protect the network and the user in order to achieve network performance objectives. The ATM Forum has also defined this set of functions as Traffic Management [42].

Recommendation I.371 describes the general objectives of traffic and congestion control functions as follows:

- ATM layer traffic control and congestion control should support a set of ATM layer QoS classes sufficient for all foreseeable services

- ATM layer traffic control and congestion control should neither rely on AAL protocols which are service specific nor on higher layers which are application specific. The control functions must remain service and application independent

- minimize network and end-system complexity.

[1]In late 1996, ITU decided to move this AAL to AAL2

Both recommendation I.371 and Traffic Management 4.0 fulfill these objectives in a slightly different way. The core of both documents is divided into three main parts, namely, the functions, the ATM Transfer capabilities or service categories and the traffic contract. Differences can be found in all three points. The goal of this section is not to give a comparison of both documents but rather to highlight the main topics of interest for this thesis.

The functions to be found in all ATM networks to control both the incoming traffic and the network congestion are (Fig. 2.16):

- Connection Acceptance Control (CAC): is defined as a set of actions taken by the network at connections setup (or connection renegotiation) in order to determine if a connection request should be accepted or rejected. It uses a traffic contract and network information on resource availability to decide.
- Usage Parameter Control/Network Parameter Control (UPC/NPC): is defined as a set of actions taken by the network to control the incoming traffic of a connection. It assures that the traffic is conforming to the traffic parameters declared in the traffic contract. The conformance definition of the incoming data is performed by the *Generic Cell Rate Algorithm* (GCRA) which may be implemented as a continuous state leaky bucket or as a virtual scheduling algorithm [42]. The main purpose of UPC and NPC is to protect the network resources from user misbehavior which may affect the QoS of other already established connections. UPC actions include discarding or tagging of cells which violate the contract established between the user and the network.
- Network Resource Management (NRM): ATM allows a logical separation of a physical transmission channel into separated Virtual Path Connections (VPC). They can be used to separate connections according to their characteristics and QoS requests.
- Priority Control or Cell Loss Priority (CLP): some applications may generate traffic flows with different cell loss priority marking. If the network treats such cell marking, it can selectively discard low priority cells to protect the high priority flows and their QoS objectives requested.
- Traffic Shaping: is defined as a mechanism that alters traffic characteristics of a cell stream in order to achieve better network efficiency while still meeting QoS objectives. It is also used to ensure conformance at a subsequent interface. In particular it may be used to smooth out the burstiness characteristics of a variable bit rate data stream.
- Fast Resource Management or ABR Flow Control: are a set of functions used to regulate traffic sources. They make use of Resource Management (RM) cells and operate on the round-trip time scale. RM cells conveys network status and allocations information to regulate the sources and to allocate network resources dynamically. ATM Block Transfer (ABT) (see below) not defined by the ATM Forum also makes use of these mechanisms.

A specific subset of these generic functions together with relevant traffic parameters as well as appropriate control functions are combined to create the *ATM Transfer Capabilities* (ATC) also called *service categories* by the ATM Forum. They are defined according to three main QoS parameters; the Cell Transfer Delay (CTD), the Cell Delay Variation (CDV) and the Cell Loss Ratio (CLR). The CTD according to ITU definition [40] measures the total transfer time of a cell. The Cell Delay Variation (CDV) can be measured in two ways; the 1 and 2-point CDV. The former describes the variability in the pattern of cell arrival events with reference to the negotiated PCR. The 1-point CDV y_k for cell k is the difference $y_k = c_k - a_k$ between the cell's reference arrival time c_k and actual arrival time a_k, where the pattern c_k is defined by:

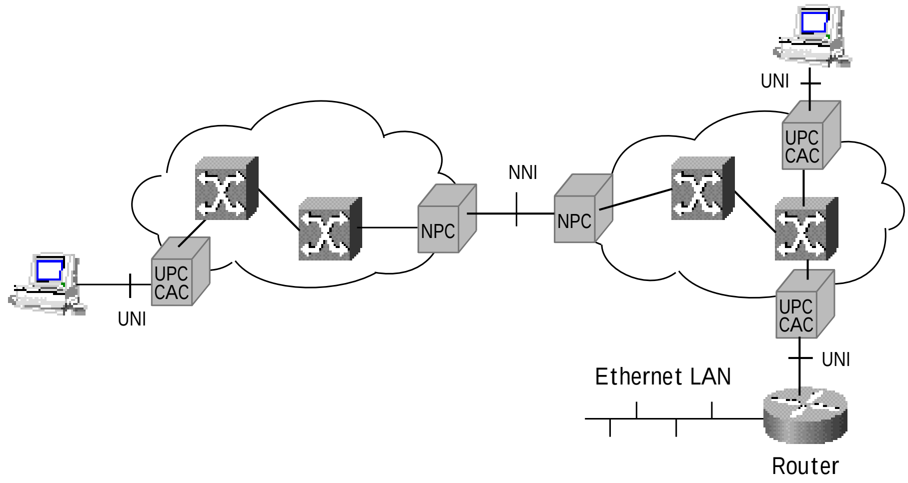


Figure 2.16: *CAC and UPC/NPC location.*

$$
\begin{aligned}
c_0 &= a_0 = 0 \\
c_k &= \begin{cases} c_k + T & \textit{when } c_k \geq a_k \\ a_k + T & \textit{otherwise.} \end{cases}
\end{aligned}
$$

The latter describes the variability in the pattern of cell arrival events at the output of a connection with reference to the pattern of corresponding events at the input. Being a_{1k} the actual arrival time at the input and a_{2k} the actual arrival time at the output, the 2-point CDV is given by:

$$v_k = a_{2k} - a_{1k} - d_{1,2},$$

where $d_{1,2}$ is the cell transfer delay measured for cell 0.

The Cell Loss Ratio (CLR) measures the ratio of total lost cells to total transmitted cells in a population of interest. The traffic parameters are needed by the CAC and UPC/NPC network functions to allocate resources, and monitor the user and the network in order to achieve the ATM layer QoS commitments related to the ATCs. The traffic parameters and the QoS requested by the user are passed to the network at connection setup via the Traffic Contract. This contract is composed of four elements namely the source traffic descriptor, the QoS class, the ATM Transfer capability and the CDV tolerance. The source traffic descriptor is the generic list of traffic parameters which can be used to characterize the traffic. These parameters are three: the Peak Cell Rate (PCR), the Sustainable Cell Rate (SCR) and the Intrinsic Burst Tolerance (IBT). The PCR, specifies an upper bound on the traffic that can be submitted on an ATM connection. It can also be described in terms of Peak Emission Interval (PEI) which describes the minimum cell interarrival time and is equal to the inverse of the PCR. The SCR specifies an upper bound on the average of conforming cells of an ATM connection over time scales which are long relative to those for which the PCR is defined. This parameter in conjunction with the Maximum Burst Size (MBS) at PCR defines the Intrinsic Burst Tolerance (IBT) as follows:

$$\tau_{IBT} = (MBS - 1)(T_{SCR} - T_{PCR}).$$

The QoS class is based on the same three parameters used to define the ATCs. The Four ATCs defined by the ITU are:

- Deterministic Bit Rate: DBR transfer capability is requested by connections that need a static amount of bandwidth available during the connection lifetime. The requested amount of bandwidth is characterized by the Peak Cell Rate (PCR). The ATM Forum equivalent service category is Constant Bit Rate (CBR) which is in fact the old name given by ITU. The ATM Forum specifies that this service is intended to support real-time applications requiring tightly constrained delay variation such as voice or video although it is not restricted to these applications.

- Statistical Bit Rate: SBR transfer capability is requested by connections which can describe in greater detail than just the PCR the traffic characteristics with SCR/IBT parameters. The ATM Forum goes far beyond this definition. They split the SBR ATC into two service categories; the Real-Time Variable Bit Rate (rt-VBR) and Non-Real-Time VBR (nrt-VBR). In their definition, rt-VBR is considered as a service category intended for real-time applications whose traffic characteristics can be described in terms of PCR, SCR and Maximum Burst Size (MBS). They specify that this service is appropriate for voice and video and that these services may support statistical multiplexing. The nrt-VBR definition is equivalent to rt-VBR with the only difference being the time constraint which is relaxed to the limit, since no delay bound is associated with this service category.

- ATM Block Transfer: ABT transfer capability is provided as a service where the ATM layer transfer characteristics are negotiated on a per-block basis. An accepted ATM Block receives the same QoS as a DBR connection with the same PCR as the one negotiated for the block. The principle of ABT is to dynamically negotiate the PCR on a per block basis. An ATM Block is a group of cells delineated by two RM cells. ABT allows two modes of transfer; ABT/DT or Delayed Transmission and ABT/IT or Immediate Transmission. As expected, the delayed transmission capability sends the block once the resource allocation has been achieved. The immediate transmission capability sends the blocks without any positive acknowledgment from the network. This may lead to loss of blocks which may be discarded if not enough resources are available at some point in the network. The ATM Forum has not defined an equivalent service.

- Available Bit Rate: ABR transfer capability is provided as a service for users that have the ability of adapting their data transfer rate according to network feedback. As a consequence, a user that adapts its traffic in accordance with the feedback information is expected to experience a low cell loss ratio. However, this mechanism neither controls the cell delay nor the cell delay variation. The traffic descriptor used for ABR is composed of a large number of parameters [42] which may fix maximum and minimum cell rates, increment and decrement cell rates, etc.

The ATM Forum has specified another service called Unspecified Bit Rate (UBR) which is in fact a *best effort* service intended for non-real-time applications. No commitments are made with respect to QoS. The UBR users do not specify any traffic parameter. Typical applications that shall make use of UBR are ftp or email.

The ATM layer QoS commitments are probabilistic only and are intended to be a first approximation of the performance the network expects to offer during a connection.

2.3 Preliminary Conclusion

If we consider what are the user and application requirements regarding real-time multimedia applications we conclude that such applications need a timely and reliable transmission in point-to-point as well as point-to-multipoint configurations. In addition the nature of the sources typically is of variable rate. The current constant rate of video applications is artificially created by modifying the image quality. If high quality audiovisual applications are the target then VBR sources are to be expected. Since VBR video has a large peak-to-mean ratio it is *economically* interesting to apply statistical multiplexing to such connections. Because statistical multiplexing *statistically* guarantees the QoS some loss may occur. Since video tolerates some, yet limited loss, it is possible to achieve a better network utilization without noticeable quality degradation.

Among the different ATC's SBR, ABR, ABT and UBR are able to handle VBR sources. ABR offers a reliable transfer but does not fulfill the timeliness required by interactive multimedia since no guarantees on CDV and CTD are offered. ABT features a per block renegotiation that could better match the variability of VBR video. ABT/DT allows to renegotiate the QoS for each block to be transferred. If accepted, the block will be transmitted with a DBR QoS. The problem with this scheme is that it is not guaranteed that the connection for the new block will be accepted generating a difficult to control end-to-end delay. If ABT/IT is used, then the probability of loss of a full block exists if the network resources are not available. UBR does not offer any guarantee and thus could hardly be used for interactive video if a high quality is expected. Finally, SBR, or more specifically rt-VBR as defined by the Forum, offers a traffic contract matching the nature of VBR data flows and provides loss and delay guarantees as required by interactive real-time multimedia applications. All ATCs but SBR could be used to transmit VBR video but their QoS delivery will depend on the network conditions.

Even if VBR video seems to be the natural choice, we have seen that among the available ATM adaptation layers none are suited for such services even though the SBR or more explicitly the rt-VBR are defined. AAL1 is able to cope with both timing and loss constraints but only if the sources deliver a constant bit rate. ITU expected this AAL to be used for CBR real-time applications which is the reason why improvements were made to reduce the extra delay caused by the FEC mechanism. The other three AALs can handle VBR sources. AAL2 is the AAL that covers all the requirements for real-time multimedia applications but by the time of writing no agreement has been reached on the specification. AAL3/4 does not provide timing functions, it is too complex and adds too much overhead albeit it is the only AAL with true multipoint functions included. The only possible choice is therefore AAL5. The ATM Forum specified the transmission of CBR encoded MPEG-2 audiovisual streams over AAL5 and foresees its extension to VBR applications also [43]. The problem with this AAL is that it is not able to reliably cover the timing and robustness requirements of multimedia due to its design suited for data transfer applications.

Clearly, no AAL is adapted to what is called to be one of the major sources of traffic in the B-ISDN, multimedia applications.

Chapter 3

Transmission of Real-Time Multimedia Streams over ATM: State-of-the-Art

3.1 Introduction

The transmission of real time multimedia streams over ATM involves a large set of topics because the interactions between the user, the coding system and the network are quite complex. Originally, video coding standards in general have not been developed with a networking perspective. Simultaneously, network standards have not been developed with a particular view on multimedia. The coders generate naturally variable bit rate data which moreover has stringent constraints in terms of delay and loss. ATM is able to handle VBR traffic. However, it may be difficult to guarantee QoS parameters. To improve this situation reliable transmission techniques may be applied which involves both the coder and the network. The QoS parameters are related to the ATM layer and are completely irrelevant to the final receiver, *the human eye and ear*. The mapping of QoS parameters, as defined at the network and perceived by the user, requires the study of the structure of the coded information and the impact of data loss onto the image which also depends on the cell loss process and the traffic profile of the source.

This chapter covers the state-of-the-art of transmission of interactive real-time multimedia data over ATM. Even though the review provides a general overview, a strong emphasis is put on the transport of MPEG-2 data. The following section is a review of the literature concerning the characterization of cell loss processes and their impact onto the user perceived QoS. Section 3.3 describes the means to reduce the impact of data loss. The ways of doing this are basically four: robust coding, error correction or reliable transmission, error concealment and rate control. We cover the first three topics the fourth being out of the scope of this work. Section 3.4 covers a particular issue of multimedia: multipoint communications. Multipoint configurations are characteristic of multimedia since they enable cooperative work, video conferencing, video broadcast, all of which are not traditional data communication configurations. The last section covers the current status of the standards in the domain of multimedia communications.

3.2 Impact of Cell Losses on Video

3.2.1 Cell Loss Characterization

One of the QoS parameters related to an ATM connection is the cell loss ratio. It is defined as the ratio of total number of lost cells to total number of transmitted cells of a connection. This parameter, a mean value, does not completely characterize the loss an application does experience. A first order statistic does not capture the cell loss dynamics. A cell loss process in the same way as traffic sources, can be characterized by other parameters such as the *burstiness*. The cell loss burstiness can dramatically modify the QoS experienced by an application. Let us consider for instance a data transfer applications using TCP/IP. If a cell loss process is bursty, then the goodput (i.e. the amount of useful information) of an application will not suffer, because the probability of having multiple cells lost in a single packet is high. Therefore, less packets will be corrupted by cell losses thus reducing the number of retransmissions. If conversely, the process is uniformly distributed, then the probability of having a single cell lost in a packet is higher than in the bursty case. The goodput will drop due to the large number of retransmissions required. Thus, leading to poor performance and to an increased connection time which, in this case, might be the determining QoS factor experienced by the user.

In the particular case of multimedia the impact that the cell loss distribution may have is different. Due to the tight timing constraints, error correction based on retransmission is in general not applicable. Therefore, cell loss will not have any impact on the duration of the call, which will remain unchanged, but will produce a quality degradation in the audiovisual data perceived by the user. Clearly, bursty losses will degrade a reduced set of frames in a more noticeable way than uniformly distributed losses which will degrade more frames and thus, the quality.

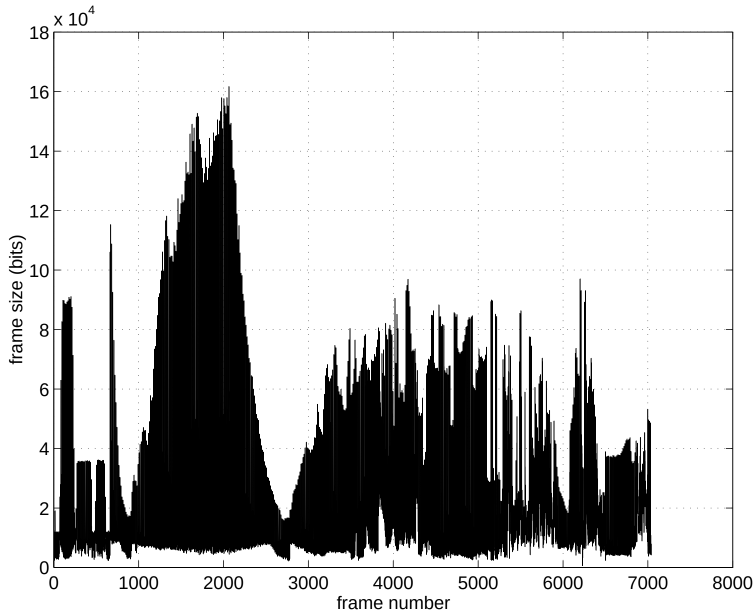


Figure 3.1: *Example of a 5 min. intraframe only MPEG trace from the movie Star Wars.*

Today, it is still unclear what the situation concerning cell losses will be. The early values provided by ITU-T (close to 10^{-9}) do not seem to match research results. In particular, it has been shown that ATM-based switching networks carrying VBR services may have non-negligible cell loss ratios [44, 45]. But it is well known that the rate of compressed video information is of highly variable and unpredictable nature (see Fig. 3.1). Basically, cell losses occur in ATM due to congestion situations that lead to buffer overflow in the network elements, especially in the switches. Two situations may generate congestion: equipment failures and multiplexing overload. The former has a likelihood estimated to be below 10^{-7} [46]. The latter will be more likely to occur if statistical multiplexing [47, 48] is used. If the operators apply peak cell rate allocation to all the connections, the probability of observing cell loss is relatively low. However, such a precautious strategy would result in an uneconomical network utilization. If, conversely, statistical multiplexing is used, then the operators will achieve a much higher resource utilization which obviously leads to increased income.

The statistical multiplexing concept is justified by the law of large numbers. The law of large numbers is founded on several theorems. Among them, the Kintchine theorem given here without proof is one of the most importants. Let us consider a collection $\{X_i; i = 1 \ to \ n\}$ of independent and identically distributed random variables such that $E\{X_i\} = \mu < +\infty$. Also consider the partial sum S_n as:

$$S_n = \sum_{k=1}^{n} x_k. \tag{3.1}$$

The Kintchine theorem states that:

$$\frac{S_n}{n} \rightarrow \mu \ as \ n \rightarrow \ +\infty \ with \ probability \ 1. \tag{3.2}$$

Restated in other words, this law says that the aggregation of a large number of variable sources will diminish the global probability to observe occurrences of aggregated bit rates that exceed a given threshold, namely the mean of the sum. Henceforth, the statistical multiplexing allows to use a transmission rate per source close to the mean rate instead of requiring a rate close to the peak. This is called Statistical Multiplexing Gain (SMG) and is often expressed as follows:

$$SMG = \frac{\sum SCR}{\sum PCR}, \tag{3.3}$$

where SCR and PCR correspond to the definitions of Sec. 2.2.4.

The larger the number of sources, the higher the gain will be. Resources are allocated based on statistical properties of the sources. Therefore, the probability exists that all the sources simultaneously transmit at their PCR. In this situation, the overall offered traffic exceeds 100% which leads to buffer overflow and cell loss.

The SCR is commonly referred to as the *effective bandwidth*. The effective bandwidth expresses the bandwidth required by an application to achieve a given QoS commonly described in terms of CLR. Several techniques to calculate the effective bandwidth are found in the literature [49, 50, 51].

Queueing theory has extensively been used to characterize cell loss ratios under different types of input traffic. A lot of work has been devoted to study the performance of ATM multiplexers to derive Cell Loss Probabilities (CLP) and queueing delays. Early works used correlated models based on Markov processes to derive performance [52, 53, 54, 55]. However, there has not been a main focus on the characterization of the cell loss process. More recent works which aim at studying the performance of video transmission over ATM make use of

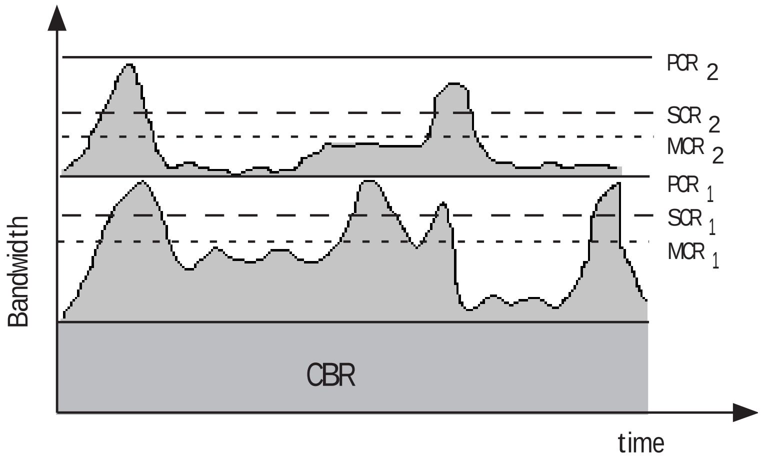


Figure 3.2: *Bandwidth allocation schemes.*

self-similar source models [56] and also available JPEG or MPEG-1 traces [24, 25] like the one shown in Fig. 3.1. The consequences that self-similarity has on the cell loss process are still not well understood. However, ATM should accommodate several types of traffic such as CBR, self-similar on-off, etc.... There is no general traffic model to accurately describe such a mix of different traffic types. As a consequence, there is no general cell loss process model since the cell loss dynamics heavily depend on the traffic profile.

Although, consensus seems to have been reached concerning two points: the first is that in case of congestion, cell losses will occur in clusters. The second is that if a single connection using relatively low bandwidth is observed, the cell loss process can be reasonably approximated by a *uniformly distributed process*. In [45], experiments show that high peak rate connections will suffer from congestion and observe high CLRs and correlated CLPs. The cell loss process of low peak rate connections can however be approximated by a uniformly distributed process because consecutive cells are sufficiently far apart so that they will see uncorrelated buffer occupancies. The same conclusion is also found in [57]. The author studies periodic traffic superimposed to background On-Off traffic and observes that the cell loss process is uniformly distributed as long as the bandwidth used by the traffic under test is less than 10% of the channel capacity. Finally, Norros *et al.* show a mathematical analysis in [58] which concludes that in a congestion period, the probability of consecutive losses remains nearly constant and close to 10^{-2} regardless of the global CLR value. The authors also show that the loss probability of a single source has not a strong dependence on the traffic characteristics. The analysis covers On-Off and gaussian sources.

There is no proof that this observation could be generalized to all types of sources. However, if it is possible to model cell loss processes by a uniformly distributed process under given conditions regardless of the source type then analytical approximations could be done.

The problem of characterizing the cell loss process is described in [59, 47, 60]. Cohen *et al.* compare four models of cell loss processes namely a uniformly distributed model, a Cohen and Heymann model [61], a On-Off model and a fourth model developed in the paper based on a Markov Modulated Process.

The uniformly distributed model, oddly called random model is the simplest loss process. This model assumes that the cell losses are independent and identically distributed with probability p. This leads to the following properties:

- the inter-cell loss times are geometrically distributed (memoryless)

- the number of loss occurrences in a time interval follows a binomial distribution.

The Cohen-Heymann (CH) model is a two-state model with a loss state, the down state, and an up state in which losses do not occur. The duration k of a loss state follows a negative binomial distribution. Given the duration k, the number of cells lost in this interval is also given by a negative binomial distribution as follows:

$$\lambda_k = \frac{0.0372}{k^{1.053}} \textit{ and } s_k = 1.097 + 0.065k. \tag{3.4}$$

To obtain the number of cell losses per source, this model relies on the assumption that the sources are independent and therefore in a multiplex of sources with the same mean traffic the losses are equiprobable. Since the sum of k independent negative binomial random variables with the same parameters λ and $\frac{s}{k}$ is a negative binomial distribution they derive that in an interval of size k, the number of cell losses per source follows a negative binomial distribution with parameters λ_k and $\frac{s_k}{k}$ as defined in Eq. 3.4.

The Gilbert model is in fact a two-state Markov chain with parameters p and q defining the transition probabilities between the loss and no-loss states such that:

$$p + q \neq 1. \tag{3.5}$$

This leads to a *burst-silence* process that is not memoryless and has geometrically distributed burst (r) and silence (s) durations

$$\left\{ \begin{array}{rcl} Prob(r) & = & q(1-q)^{r-1} \\ E(r) & = & \frac{1}{q} \end{array} \right.$$
$$\left\{ \begin{array}{rcl} Prob(s) & = & p(1-p)^{s-1} \\ E(s) & = & \frac{1}{p} \end{array} \right.$$

The condition of Eq. 3.5 leads to correlated losses. In this case the Autocovariance is defined by:

$$Autocov = \frac{pq}{(p+q)^2}(1-p-q)^{|k|}. \tag{3.6}$$

The new model developed by the authors combines properties of the CH model and the Gilbert model. The authors propose that in the loss state, the number of cells lost during a burst of size k follows a negative binomial distribution such as in the CH model defined by:

$$P(loss_k) = \frac{1.097 + 0.065^k}{0.372} k^{1.053}. \tag{3.7}$$

The model obtained is a clustered cell loss process. The authors use the same assumptions than for the CH model, iid sources with same mean parameters. The cell losses experienced by a single source are then derived by assuming that the cell losses can be distributed among the connections which brings the final cell loss process, per source, to an approximation of a uniformly distributed model.

In [60] spectral analysis techniques are applied to study correlations in the cell loss interoccurrence process. By assuming independent and identically distributed interoccurrences, assumption not necessarily found in real networks, the authors show that the cell loss process is of type *General Independent* (GI) which corroborates the measurements done via simulation. This study however cannot be fully generalized to all kinds of input traffic.

A different aspect of the problem is tackled in [44, 45, 62, 63] where the impact of cell losses on frame corruption is studied. In [44] it is shown that frame loss can unacceptably grow due to cell losses in ATM networks when some burstiness on the frame time scale occurs. The work developed in [45] also studies the link between cell and frame loss based on peak cell rate. The authors conclude that at a frame level, bursty losses cause less damage because the cells lost have a higher probability to belong to a single frame. The impact that this has on retransmission mechanisms is studied in [63] and the author's conclusion is that frame based retransmission techniques are more effective than cell based ones assuming a bursty loss model. Conversely, the uniformly distributed loss process is the worst cell loss distribution in terms of QoS impact.

An interesting conclusion found in these papers is that the significance the CLR has as a QoS parameter for frame based applications and in particular for real-time multimedia is doubtful. In fact QoS should be specified *at least* at the AAL level to take into account the framed nature of the data.

This raises one of the problems of cell loss characterization related to multimedia applications. Whether the studies focus on networking issues or on user perceived QoS, the results can be analyzed differently and conclusions can be very different. Acceptable error rates for video applications depend on such criteria as the quality of the original data, duration of the audiovisual material and even on the location of the error within the image itself (known as the focus of attention). Also the utilization of error concealment techniques may influence the perceived QoS. How all these criteria could be mapped to lower layer QoS parameters is a matter of current research.

3.2.2 Mapping Cell Loss to User Perceived Quality

As already observed, the mapping of ATM layer QoS values to frame or AAL level QoS is not straightforward. This same problem holds for higher layers and in particular for multimedia applications. The impact that the network QoS parameters have on an application level QoS or even on the user perceived quality is very difficult to map.

The description of the different compression algorithms of Chap. 2 shows that all the current compression algorithms structure the information in a hierarchical way. A syntax based on a set of headers is required for the decoder to interpret the received data. If syntactic information is missing the decoder is unable to interpret the encapsulated information. As an example, in an MPEG-2 video stream, data loss reduces quality depending strongly on the type of the lost information. Losses in syntactic data, such as headers and system information, affect the quality differently than losses of semantic data such as pure video information. Furthermore, the quality reduction depends also on the location of the lost semantic data due to the predictive structure of an MPEG-2 video coded stream.

Let us consider Fig. 3.3 showing how network losses map onto visual information losses in different types of MPEG-2 pictures. Data loss spreads within a single picture up to the next resynchronization point (e.g. slice headers) mainly due to the use of variable length coding, run length coding and differential coding. This is referred to as *spatial propagation* and may damage any type of picture. When loss occurs in a reference picture (intra-coded or predictive frame), it will remain until the next intra-coded picture is received. This causes the errors to propagate across several non intra-coded pictures until the end of the current GOP. This is known as *temporal propagation* and is due to inter-frame predictions.

The impact that the loss of syntactic data may have is in general more important and more difficult to recover than the loss of semantic information. This data loss may induce frame loss in the decoded sequence. Indeed, when a frame header (a few syntactic bytes before each frame in the bitstream) is lost, the entire corresponding frame is skipped because the decoder is not able to detect the beginning of the frame. If the skipped frame corresponds

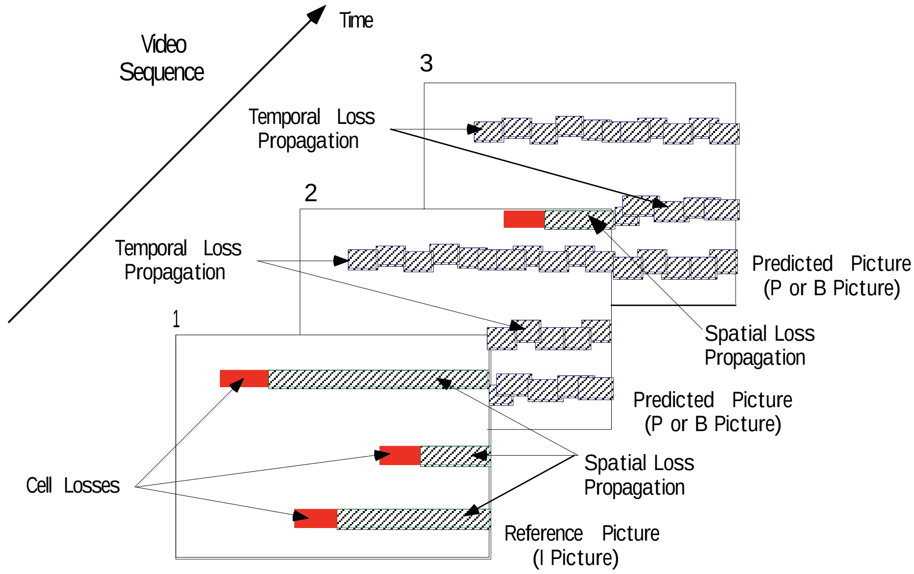


Figure 3.3: *Data loss propagation in an MPEG-2 decoded video sequence.*

to a predictive picture (I or P), it may strongly reduce the perceptual quality due to the predictive structure of the MPEG-2 video stream.

The problem is actually that when a header is lost, in general, the whole information it carries is skipped. Some headers are thus more crucial than others. For instance, sequence headers, predictive (I or P) picture headers, PES headers, slice headers in intra-coded pictures can be considered as essential in comparison to slice headers in B pictures.

Another reason which makes mapping of the network QoS onto user QoS difficult is the lack of a video quality metric that behaves according to the user. Traditionally, the quality metric used for audiovisual signals is the Peak Signal to Noise Ratio (PSNR). It is a quantitative measure of the distortion of an image compared to the original defined as:

$$PSNR(dB) = 10 \log_{10} \frac{\sum_{i=1}^{N_p} o_i^2(n)}{\sum_{i=1}^{N_p} (o_i(n) - d_i(n))^2}, \tag{3.8}$$

where $o_i(n)$ and $d_i(n)$ are the luminance values for the i'th pixel of the n'th original and encoded frames, respectively and N_p is the number of pixels in a frame. It has been proved in [64] that the PSNR has no correlation with the user perception of video quality. The linear relationship between the encoding bit rate and the PSNR metric shown in Fig. 3.4is in contrast with the human inability to perceive a quality improvement beyond a certain encoding bit rate.

An attempt to map network QoS to higher layer QoS is to be found in [65]. The metric called *Glitch* tries to capture the impact of ATM cell losses at a display level for MPEG-2 applications. The problem with this metric is that it does not give any real information of what is perceived by the user. Indeed, a glitch is defined as the interval started by an image partially displayed and ended by an image correctly displayed. The glitch duration and glitch rate are the main metrics. Unfortunately, at a user level, the perception of a glitch depends on the duration but also on the portion of the image that is damaged. This third metric is not considered by the Glitch metric.

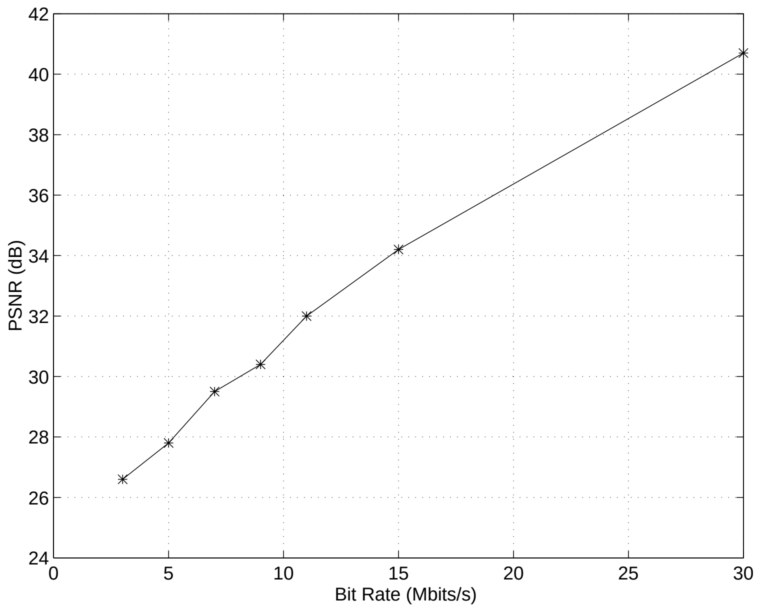


Figure 3.4: *PSNR as a function of the encoding bit rate.*

Recent research has addressed the issue of video quality assessment by means of human correlated metrics. One of the first quality metrics, the $\hat{s}$ (SHAT), was developed at the Institute for Telecommunication Science (ITS) in Colorado [66]. The quantitative measure is based upon two quantities, namely, spatial and temporal information, *SI* and *TI* respectively, which tries to map subjective evaluations. The spatial information is defined as:

$$SI(F_n) = STD_{space}\{Sobel[F_n]\}, \tag{3.9}$$

where STD_{space} is the standard deviation operator over the horizontal and vertical spatial dimensions in a frame, and *Sobel* is the Sobel filtering operator, a high pass filter, used for the edge detection.

The temporal information is based upon the motion difference image ΔF_n calculated as the pixel difference among successive frames. The temporal information is defined as:

$$TI(F_n) = STD_{space}[\Delta F_n]. \tag{3.10}$$

The SHAT metric is a linear combination of three quality impairment measures namely m_1, m_2 and m_3, defined upon the SI and TI metrics. The m_1 metric is a measure of spatial distortion and is obtained upon the *SI* information as follows:

$$m_1 = RMS_{time}(5.81 \times |fracSI[O_n] - SI[D_n]SI[O_n]|),$$

where O_n is the n^{th} frame of the original video sequence, D_n corresponds to the n^{th} frame of the degraded sequence and RMS denotes the root mean square function. The subscript time means that the function is performed over time.

The remaining two measures m_2 and m_3 are temporal distortion metrics and are given by:

$$m_2 = f_{time}[0.108MAX\{(TI[O_n] - TI[D_n]), 0\}],$$

where

$$f_{time} = STD_{time}\{CONV(x_t, [-1, 2, -1])\}.$$

STD_{time} is the standard deviation across time. $CONV$ is the convolution operator.

$$m_3 = MAX_{time}\{4.23LOG_{10}(fracTI[D_n]TI[O_n])\}.$$

Finally, the SHAT metric is given by the following linear combination:

$$\hat{s} = 4.77 - 0.992m_1 - 0.272m_2 - 0.356m_3.$$

It is however shown in [64] that this metric over estimates the quality in the low bit rate encoding range of MPEG-2 (see Fig. 3.5).

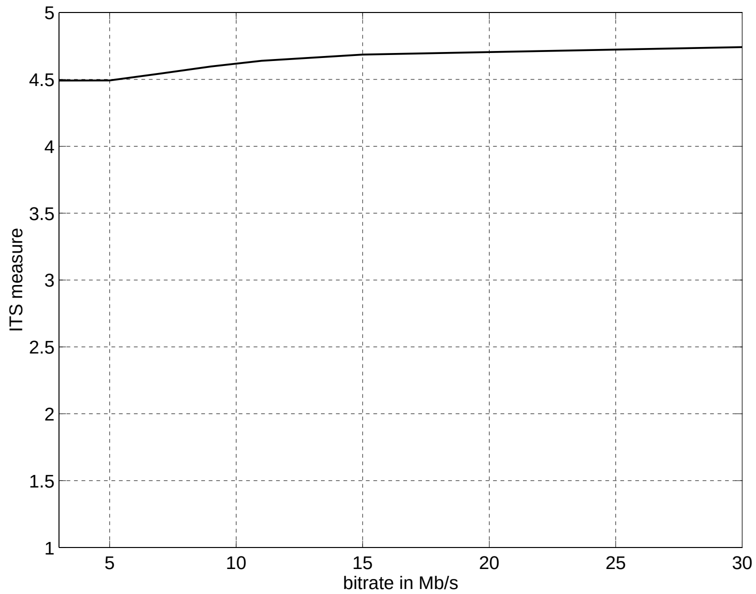


Figure 3.5: *SHAT as a function of the encoding bit rate.*

Recently, several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of the Human Visual System [67]. A spatio-temporal model of human vision has been developed for the assessment of video coding quality [68, 64]. The model is based on the following properties of human vision:

- The responses of the neurons in the primary visual cortex are band limited. The human visual system has a collection of mechanisms or detectors (termed "channels") that mediate perception. A channel is characterized by a localization in spatial frequency, spatial orientation and temporal frequency. The responses of the channels are simulated by a three-dimensional filter bank.
- In a first approximation, the channels can be considered to be independent. Perception can thus be predicted channel by channel without interaction.
- Human sensitivity-to-contrast is a function of both frequency and orientation. The *contrast sensitivity function* (CSF) quantizes this phenomenon by specifying the detection threshold for a stimulus as a function of frequency.

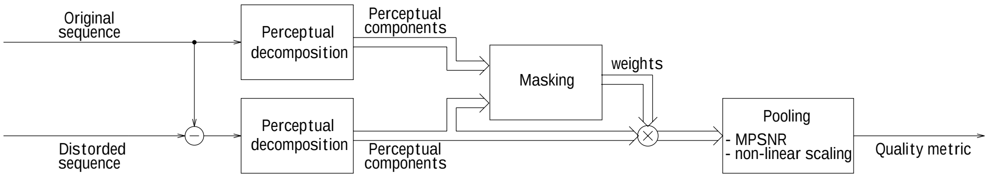


Figure 3.6: *Moving Pictures Quality Metric (MPQM) block diagram.*

- Visual masking accounts for inter-stimuli interferences. The presence of a background stimulus modifies the perception of a foreground stimulus. Masking corresponds to a modification of the detection threshold of the foreground according to the local contrast of the background.

The vision model described in [68] has been used to build a computational quality metric for moving pictures [64] which proved to behave consistently with human judgments. Basically, the metric, termed Moving Pictures Quality Metric (MPQM), first decomposes an original sequence and a distorted version of it into perceptual components by a Gabor filter bank. Indeed, the profile of the channels is very close to Gabor functions. A channel-based distortion measure is then computed accounting for contrast sensitivity and masking. Finally, the data is pooled over the channels to compute the quality rating which is then scaled from 1 to 5 as described in Tbl. 3.2.2 [69] (see Fig. 3.6).

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Table 3.1: *Quality scale that is often used for subjective testing in the engineering community.*

Recently, an improved metric, termed Normalized Video Fidelity Metric (NVFM), is based on a finer modeling of vision and has been introduced in [70]. This new metric adds a modeling of the saturation characteristic of the cortical cells' responses and a modeling of inter-channel masking. It is an extension of a still-picture model developed by Teo&Heeger [71].

A limitation of vision models is that perception above what is called the contrast threshold is not known. In vision science, the contrast threshold CT is defined as the necessary contrast of a stimulus at a given frequency (spatial and temporal), orientation and color to provoke a neuronal response,

$$sensitivity = \frac{1}{CT},$$

where CT is a function of the CSF.
Then , the suprathreshold which defines the limit of current models is given by:

$$suprathreshold > CT. \tag{3.11}$$

Experiments have shown that the suprathreshold is bigger than 2 CT.

A study of quality assessment of the MPQM metric is to be found in [64]. A set of test video sequences have been encoded at different bit rates. The MPQM quality has then been evaluated. Fig. 3.7 presented in [64] shows that the perceptual quality saturates at high bit rates. This leads to the conclusion that increasing the bit rate may thus result, at some point, in a waste of bandwidth since the end user does not perceive an improvement in quality anymore.

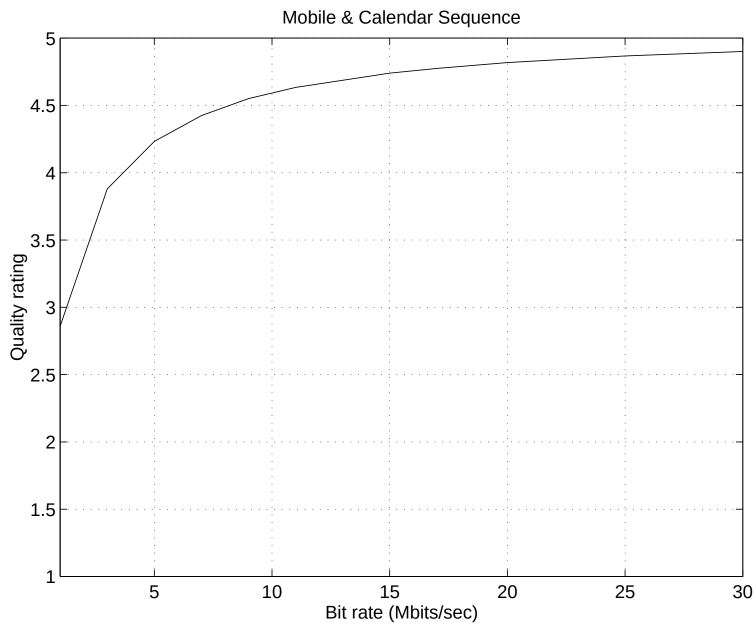


Figure 3.7: *MPQM quality assessment for the Mobile & Calendar sequence as a function of the bit rate.*

3.3 Reliable Transmission

Reliable transmission covers four aspects:

- robust coding which aims at reducing the impact of data loss by modifying the coding schemes or use special features to better fit to packet networks
- error correction which is mainly based on the utilization of forward error correction techniques
- error concealment which targets the decoder and aims at masking the impact of cell losses onto the image
- rate control which tries to control or adapt the application bit rate by different means.

Among others, these are network feedback based bit rate control, shaping techniques and joint source channel coding. We cover in this section the state of the art on the first three topics, the fourth being out of the scope of this work.

3.3.1 Robust Coding

The main goal of robust coding is to reduce the impact of data loss at the decoder. Basically, there are two ways of improving the coding, namely, source coding and channel coding. Source coding which deals with the original data aims at developing encoding mechanisms to deliver compressed data as robust as possible. On the other hand, channel coding which deals with the already coded data tries to find techniques, which basically are data reorganization and packetization, to reduce the impact of data loss onto video. Some work has been done on joint source/channel coding techniques which generally link the network and the codec. However, since joint coding techniques are based on feedback they are difficult to apply to real-time applications and are therefore not discussed here. The literature mostly covers three compression standards; H.261 [11], MPEG-1 [10] and MPEG-2 [17, 16]. We develop in this section channel coding only since source coding is not a networking issue and is out of the scope of this work.

The early works on robust coding targeted H.261 codecs because of their availability and advanced standardization. One method used to improve robustness of transmission consists of using a *layered coding* technique as described in Chap. 2. Compression techniques generally allow to generate multiple bit streams. One of the bit streams contains the essential information for the decoder, the *base layer*. The remaining layers provide enhancements to the image quality and are known as *enhancement layers*. This layers do not contain any essential information. Different techniques have been proposed to take advantage of the scalability of the video coders. An adaptive two-layer H.261 based codec is proposed in [72]. The enhancement layer is a finer requantization of the residual quantization distortions of the base layer. Another two-layered H.261 codec is proposed in [73]. It uses a data partitioning method to define the layers. The base layer contains the DC coefficients and part of the AC coefficients while the enhancement layer contains only AC coefficients. The number of AC coefficients in the base layer is calculated to always have a PSNR above a fixed threshold. The enhancement layer is used for error concealment at the decoder. The results show that partitioning the data under the assumption that the base layer has a lower CLR than the enhancement layer gives better performance than the transmission of a single layer under the same CLR as the base layer. A two-layered subband coding for H.261 is proposed in [74]. The different subbands are mapped to different layers. The author also proposes a specific ATM cell format.

The packing of H.261 video data into ATM cells is studied in [75]. A close packing scheme, in which the information is packed continuously into a cell payload until it is full, is proposed. In this case, some macroblocks may be split between two cells. Conversely, the loose packing fills a cell payload with an integral number of macroblocks. To avoid error propagation, it is also proposed to add extra information such as a cell sequence number and the absolute addressing of the first complete macroblock in the cell. Loose packing adds 8 to 9% overhead but reduces the number of affected macroblocks.

The most addressed compression technique has definitely been MPEG-2. Its position as the standard for multimedia has raised a large interest in the scientific and technical community. Also, the standard, even if complete, leaves a large place to interpretation in several topics. It provides four types of scalability to produce layered video bit streams which allows different profiles and levels that could be used in many ways including cell loss resilience [76].

One of the basic starting points to robust coding is the syntactic relevance of each of the three I, P and B pictures. Clearly the impact that loss has in I pictures is the highest since the temporal propagation is also the longest and largely contributes to the image degradation due to the spatial propagation of the error within the I frame itself. On the other side, B pictures are the less relevant since they are not used as a prediction reference and therefore do not generate temporal propagation of errors. This fact is used in [77] where the authors propose a coding mechanism that uses reduced slice sizes in reference pictures, either I and P, and longer slice sizes in B pictures. This technique is allowed by the standard which does not fix any slice size. The overhead generated by the extra slice headers is compensated by the larger slices of the B pictures. The study achieves a 1 to 1.5 dB improvement over a standard "single slice size" coding technique. However, the authors do not consider the MPEG-2 system layer which encapsulates the MPEG-2 data into fixed packets and clearly modifies the impact that cell losses have onto the video stream.

Leditschke *et al.* also take advantage of the possibility of having variable slice sizes in MPEG streams. In [78] a proposal to adapt the size of a slice to match the size of a cell is described. The authors also propose a pre/post processor that rearranges the macroblock information given that the most important information of all macroblocks in a slice is located towards the start of the slice (e.g. DC coefficients, motion vectors). The rearrangement is not standard and therefore offers a non-compatible solution with standard MPEG equipment.

Pancha *et al.* also propose a two-layered VBR MPEG coder[79]. The authors propose a data partitioning scheme with a high priority flow which contains all the header information and DC coefficients and part of the AC coefficients. The proportion of AC values depends on a priority control function. The remaining AC coefficients are sent into the low priority flow.

Based on the same adaptive mechanism, [80] proposes to combine the monitoring of a leaky bucket to detect bursts and change the compression rate to reduce the burstiness of the high priority flow. By achieving this, the high priority flow is expected to be conforming to the UPC parameters and would not suffer from data loss.

The syntactic relevance of the different types of pictures is also used in [81]. The authors propose to separate the data into three flows composed exclusively of pictures of the same type. An advantage of this method is that the burstiness of each flow is reduced compared to the burstiness of the single MPEG flow. Uniformly distributed losses are applied to each of the flows separately and the obtained PSNR is measured. The best quality is obtained when the B flow is damaged and the worst when the I flow is damaged. The quality obtained by damaging the standard MPEG stream gives a slightly better quality than the I damaged flow. These results suggest that different QoS target value could be applied to the separated flows, the better being applied to the I flow and the worst to the B flow. However, the synchronization problems that this technique could generate are not discussed in this contribution.

Some of the proposals clearly improve the robustness of video. By considering the network as a lossy channel, they try to put as many error correction and concealment mechanisms at the application level as possible. This sometimes makes network functions redundant and adds an extra overhead. A combined approach between the network and codec that better takes into account and exploits the available network functions could be more efficient.

3.3.2 Error Correction

3.3.2.1 Forward Error Correction

Traditionally, the error correction techniques have been applied first in radio channels and later in data networks. The applications that make use of these techniques are file transfer applications because no loss is tolerated. Since they do not have any particular timing requirements, the error correction techniques used are based on retransmission mechanisms and are generally performed by the transport layer (e.g. TCP). Closed-loop techniques such as *Automatic repeat request* (ARQ) are retransmission mechanisms which require that both end-systems exchange status information about each of the transmitted packets. Therefore, a delay of at least one round-trip time is added when error correction is required. Using retransmission means that the receiver needs only to check if the received data is correct. This is done with Cyclic Redundancy Codes (CRC) or parity codes. Performing error detection only at the receiver has the advantage of generating less overhead than correcting codes and also less computational effort to calculate the codes. However, these techniques can hardly be used for time constrained applications. Instead, *Forward Error Correction* (FEC) techniques can be applied. Because they recover from errors at the receiver and do not add an extra delay due to the retransmission itself, they avoid the shortcomings of ARQ.

FEC techniques have been developed for transmission on unreliable channels (e.g. satellite links) in which the timing constraints as well as the bandwidth cost avoid the use of retransmission. In high bandwidth-delay product networks, where the ratio of packet transmission time to the propagation delay is large, FEC also presents advantages compared to retransmission schemes. FEC techniques are commonly used in compact disks and hard disk technologies. However, the major advantage of FEC related to multimedia applications is its ability to handle *isochronous* data flows.

The increasing deployment of ATM which has a high bandwidth-delay product has raised the question of the utilization of FEC. The research community has devoted very much work on FEC for ATM. The main topics that have been tackled are:

- the error correcting codes to be used in ATM
- the gain of using FEC for both multimedia and data communications
- the proper place of FEC in the protocol stack. Whether FEC has to be placed at the transport layer the AAL or physical is still under debate.

The are two types of codes in common use today, block codes and convolutional codes. A block code encoder, takes a set of k input symbols represented by the $k-tuple$ $\mathbf{u} = (u_1, u_2, \ldots, u_k)$ called a *message* and generates a larger set of n output symbols. The $n-tuple$ $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is called a *code word*. This block code is referred to as a (n, k) block code. The ratio of the number of input to output symbols in a data packet is called the code rate R and is expressed as follows:

$$R = \frac{k}{n} < 1. \tag{3.12}$$

The encoder for a convolutional code also accepts *k-symbol* blocks of information and produces an encoded sequence or code word $\mathbf{v}$ of *n-symbol* blocks. However, each encoded block depends not only on the current input message but also on *m previous* message blocks. The encoder is said to have a *memory of order m*.

To understand the operation of many correcting codes it is necessary to have a basic knowledge in the area of number theory. In fact, the commonly used correcting codes are defined over algebraic systems called *fields*, and in particular *finite* or *Galois* fields.

A finite field is a finite set F of elements on which two binary operations called addition, "+" and multiplication "." are defined such that the result of a linear combination of two elements in the field is another element in the field. The set F together with the two binary operations is a field if the following conditions are satisfied:

1. F is commutative under addition. The identity element with respect to addition is called *zero*
2. the set of nonzero elements in F is commutative under multiplication. The identity element with respect to multiplication is called *unit*

+	0	1
0	0	1
1	1	0

.	0	1
0	0	0
1	0	1

Table 3.2: *Operations in GF(2)*

3. multiplication is distributive over addition.

The French mathematician Galois proved that finite fields can only be constructed if the field size is a prime number or a power of a prime number. A Galois field of order p is denoted as $GF(p)$ and is constructed using integer modulo p arithmetic. As an example, the binary field or $GF(2)$ and the associated addition and multiplication tables are shown in Tbl. 3.2. In fact, for any positive integer m, it is possible to extend the prime field $GF(p)$ to a field of p^m elements which is called an extension field of $GF(p)$ and is denoted by $GF(p^m)$.

For a field of size q^c, q being prime, the nonzero elements of the field consist of the roots of the polynomial that satisfy the following equation:

$$X^{q^c-1} + 1 = 0. \tag{3.13}$$

Consider as an example $GF(2^3)$, then Eq. 3.13 may be factored as follows:

$$\begin{aligned} X^7 + 1 &= (X^3 + X + 1).(X^3 + X^2 + 1).(X + 1) \\ &= 0. \end{aligned}$$

The irreducible polynomials of degree c are said to be primitive polynomials and are used to construct the field using arithmetic that is modulo some primitive polynomial. $GF(2^3)$ for instance can be constructed using:

$$\begin{aligned} (X^3 + X + 1) &= 0 \\ (X^3 + X^2 + 1) &= 0. \end{aligned}$$

The construction of Galois Fields is not a straightforward operation. Tables and algorithms to construct Galois Fields can be found in [82]. The next two paragraphs are an introduction to both Reed-Solomon correcting ncodes and Reed-Solomon erasure codes.

3.3.2.1.1 Reed-Solomon Codes

Reed-Solomon codes (RSC) are a subclass of the Bose, Chaudhuri and Hocquenghem (BCH) codes which form a large class of random error-correcting codes. Reed-Solomon codes are multilevel (i.e. codes dealing with non-binary or *q-ary* symbols) codes. A t-error-correcting RS code (n, k) over $GF(q)$ has the following properties:

- block length: $n = q - 1$
- number of parity-check symbols: $n - k = 2t$
- minimum distance: $d_{min} = 2t + 1$.

Lets consider a RS code over $GF(2^m)$ where $q = 2^m$. Let α be a primitive element in $GF(2^m)$. The generator polynomial of a t-error-correcting RS code of length 2^{m-1} is:

$$
\begin{aligned}
g(X) &= (X+\alpha)(X+\alpha^2)\ \ldots\ (X+\alpha^{2t}) &(3.14)\\
&= g_0 + g_1X + g_2X^2 +\ \ldots\ + g_{2t-1}X^{2t-1} + X^{2t}. &(3.15)
\end{aligned}
$$

The values $\alpha^2,\ \ldots,\ \alpha^{2t}$ are the consecutive roots of $g(X)$. The code generated by $g(X)$ is an $(n, n-2t)$ cyclic code. Let

$$
\mathbf{a}(X) = a_0 + a_1X + \ldots + a_{k-1}X^{k-1}, \tag{3.16}
$$

be the message to be encoded, where $k = n - 2t$. The $2t$ parity symbols are the coefficients generated by the polynomial division of the message $X^{2t}\mathbf{v}(X)$ by $g(X)$.

Lets consider the coefficients v_i of the message $\mathbf{v}$ to be transmitted. Then $\mathbf{v}$ can be expressed by:

$$
\mathbf{v}(X) = v_0 + v_1X + \ldots + v_{n-1}X^{n-1}. \tag{3.17}
$$

The received vector $\mathbf{r}$(X) can be expressed by the following polynomial:

$$
\mathbf{r}(X) = r_0 + r_1X + \ldots + r_{n-1}X^{n-1}. \tag{3.18}
$$

The errors can then be represented as a linear combination of the received word $\mathbf{r}(x)$ and the transmitted code word $\mathbf{v}(x)$:

$$
\begin{aligned}
\mathbf{e}(x) &= \mathbf{r}(x) - \mathbf{v}(x) &(3.19)\\
\mathbf{e}(x) &= e_0 + e_1x^1 + e_2x^2 + \ldots + e_{n-1}x^{n-1}. &(3.20)
\end{aligned}
$$

Assume an errored code word with u errors at locations $X^{j_1}, X^{j_2}, \ldots, X^{j_u}$ under the condition that $j_u \leq n-1$. Then one can rewrite $\mathbf{e}(X)$ as follows:

$$
\mathbf{e}(x) = e_{j_1}X^{j_1} + e_{j_2}X^{j_2} + \ldots + e_{j_u}X^{j_u}.
$$

To correct the received code word it is necessary to determine the error locations X^{j_i} as well as the values e_{j_i}.

In order to correct the errors, the polynomial $\mathbf{r}(x)$, the received code word, is evaluated at each of the roots α^i, giving *2t* equations that produce the syndrome. In matrix form:

$$
S = E.R, \tag{3.21}
$$

where S is the syndrome vector, E, the error vector and R the root matrix:

$$
\begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_{2t} \end{bmatrix} = [e_{i_1}, e_{i_2}, \ldots, e_{i_u}] \begin{bmatrix} \alpha^{i_1} & \alpha_{i_2} & \ldots & \alpha_{i_u} \\ \alpha^{2i_1} & \alpha^{2i_2} & \ldots & \alpha^{2i_u} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{2ti_1} & \alpha^{2ti_2} & \ldots & \alpha^{2ti_u} \end{bmatrix} \tag{3.22}
$$

Solving Eq 3.22 is not possible until the e_{i_j} values are known. To locate the position of the errors it is necessary to create an *error-locator* polynomial:

$$
\begin{aligned}
\sigma(X) &= (1+\alpha^{j_1}X)(1+\alpha^{j_2}X) + \ldots + (1+\alpha^{j_u}X) &(3.23)\\
&= \sigma_0 + \sigma_1X + \sigma_2X^2 + \ldots + \sigma_uX^u. &(3.24)
\end{aligned}
$$

The roots of $\sigma(X)$ are $(\alpha^{j_i})^{-1}$, which are the inverses of the error locator numbers. This equation can be solved using iterative algorithms such as Berlekamp's. Once $\sigma(X)$ is found, the e_{i_j} values could be determined and Eq 3.19 solved.

It is possible, given an (n, k) cyclic code, to achieve a $(\lambda n, \lambda k)$ cyclic code by *interleaving*. This is done by arranging λ code vectors into m rows of a rectangular array and then transmitting them columnwise as depicted in Fig. 3.8. The parameter λ is referred to as the interleaving degree. As Fig. 3.8 shows, if a burst of errors causes consecutive symbols to be corrupted, the errors will be spread out as single errors within all the codewords making them correctable. Interleaving decorrelates the bursts of errors to allow recovery with simple correcting codes. in fact, if the original code used corrects a single burst of length l then the interleaved code will correct any single burst of length λl or less. However, due to the necessary matrix configuration, interleaving invariably introduces delay at both the sender and the receiver sides proportional to the size of the interleaver as well as to the data rate at which the interleaver is filled.

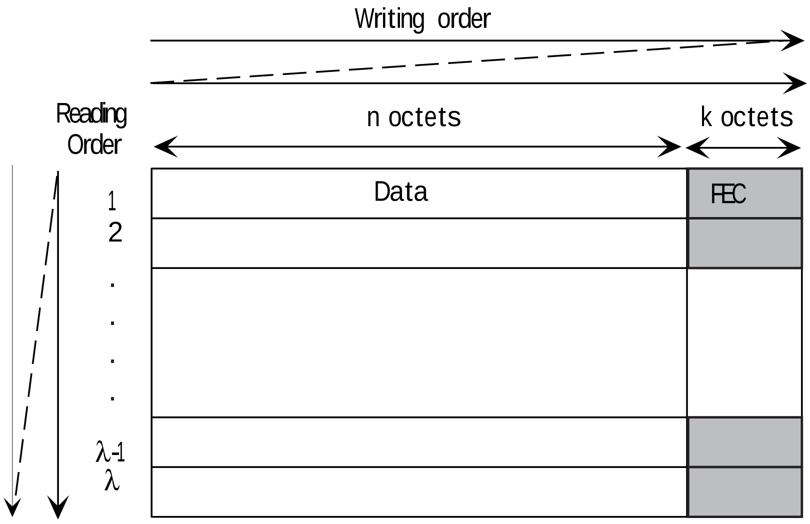


Figure 3.8: *Octet based interleaver.*

3.3.2.1.2 Burst Erasure Codes

Coding theory defines an error as a corrupted bit, or symbol, with an unknown value in an unknown location whereas an erasure is defined as a corrupted bit, or symbol, with an unknown value in an known location. In ATM networks, erasures which are the most common errors have a known size and occur only at cell boundaries. When the error location is known, the error correcting power of the code is doubled. The redundancy codes generally used for FEC are the *Reed-Solomon Codes* (RSC). RSCs are able to detect bit errors as well as erasures (data loss). This correcting power was necessary for older networks with relatively high *Bit Error Rates* (BER). However, in fiber based networks, BER are estimated to be close to 10^{-9} which is largely below the expected cell loss ratios. Therefore, the utility of such sophisticated techniques is questionable. Based on this observation, a new set of error correcting codes has been developed. The *burst erasure codes* (RSE) described in [83] are simpler to implement but are only able to correct erasures. They rely on the fact that the basic transmission unit is the cell. So, erasures are of known size and are located in known boundaries (cell boundaries). Bit errors are still detected by this mechanism but are not corrected.

The Burst Erasure correcting code is based on the described RSC codes and produces identical codewords with the same parameters. This means that an RSE code is able to correct an errored message based only on Eq. 3.22. The set of equations 3.23 and 3.24 required to locate the position of the errors is not necessary since the position is assumed to be known. This largely reduces the complexity of the FEC decoder since a single set of equations has to be solved. The counterpart to this complexity reduction is an increase in the overhead. Because, RSE codes assume the position to be known, it is implicitly assumed that a sequence number is inserted in the cells. This is relatively limiting since only AAL1 implements this feature.

3.3.2.2 AAL Level Forward Error Correction

The efficiency of FEC has been and is still questioned due to the unclear modeling of cell losses. In [84], an exhaustive study of the efficiency of RSE based FEC techniques is presented. Different mixes of traffic were used to simulate different traffic and cell loss profiles. The main conclusion of this paper is that the efficiency of FEC depends on the cell loss process. While a bursty cell loss process reduces the efficiency of FEC recovery, a uniformly distributed loss is the best case for FEC recovery. As already discussed in Sec. 3.2.1, the uniformly distributed loss model seems to be accurate for single connections while the burst loss model depicts the behavior of a global process. FEC techniques should therefore be efficient for single connections.

The work presented in [85] studies the impact that the FEC block size has on FEC recovery efficiency in the particular case of VBR video. The problem of this study is that this simulation scenario is composed only of video sources. Moreover, the authors apply a Gilbert model (see 3.2.1) which generates a correlated cell loss process. Considering this scenario, the results show as expected that the larger the FEC block size, the better the performance.

ITU-T included a (124, 128) RS code based FEC technique in AAL1 (see Sec. 2.2.3.1) targeted for circuit emulation services (e.g. telephony) and audiovisual applications such as video on demand which have stringent timing constraints. The problem of AAL1's FEC scheme is that it adds a significant delay due to the size of the interleaver. Moreover, using RSC adds a computational complexity which is reflected on Network Interface Card (NIC) street prices. Considering that, as mentioned earlier, the expected BERs are low compared to cell losses, the complexity added by the octet granularity of the algorithm is not necessary anymore.

Several papers have studied the efficiency of AAL1 for video transmission. Rasheed *et al.* [86] argue that AAL5 being a short term solution AAL1 will be adopted later and proposes the utilization of the User-to-User bit of the ATM cells to delineate the matrix. Simulation results show also that the utilization of an interleaving mechanism worsens the packet error ratio when cell recovery is not possible since the errors propagate in all the packets of the FEC block, thus corrupting the full block.

A modified version of AAL1 for the transport of video has also been proposed in [87]. The main modification consists in changing the interleaver size and the position of the packets in the matrix to reduce the delay due to the accumulation of packets in the interleaver matrix.

Concerning the transmission of VBR traffic, the utilization of interleaving poses a problem in particular for real-time services. Since an interleaver has to be filled up prior to transmission, the data packets arriving at a variable rate will experience a variable delay per packet which could be very difficult to handle at the receiver side and is therefore, a major drawback for interactive services. Still, the majority of the literature proposes variations of the AAL1 scheme.

An early work targeting the unspecified AAL2 proposed to use a packet interleaver combined with a time out mechanism [88]. The authors observe that the fixed size matrix scheme may generate too much delay for low bit rates as it could happen in low activity scenes. When the time out occurs, the *partially filled* matrix is sent. The major drawback of this technique is that to generate the redundancy cells, the matrix is filled with dummy data which is later sent. This generates a large overhead that will increase the mean cell rate of the source.

The idea of *dynamic interleaving* was presented again in ATM Forum contributions [89, 90, 91] which proposed an AAL2 for class B services as an enhanced AAL1.

In [92] the proposal is to have RS codes adapted to the applications delay constraints. Different combinations of RS codes which lead to different matrix sizes are studied. One of the conclusions is that interleaving can be avoided in several cases to reduce delay without loss of FEC correction efficiency. A mechanism to change the source rate when FEC is applied to maintain a fixed overall transmission rate is proposed by the same author in [93].

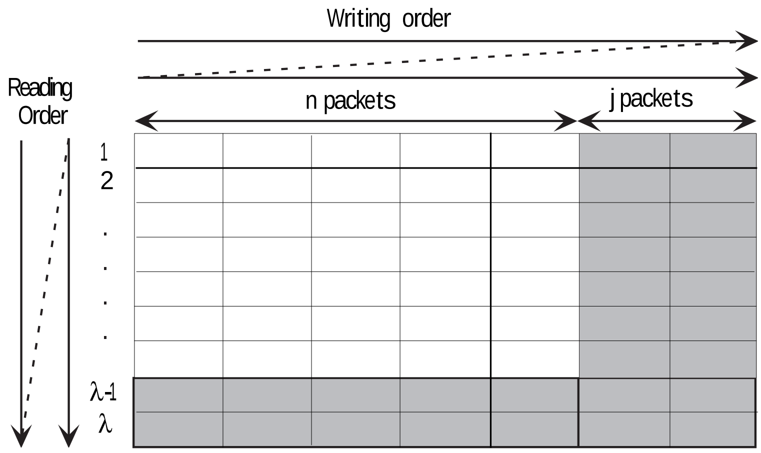


Figure 3.9: *Cell based interleaver.*

A different approach is described in [94] where a per Virtual Path (VP) error recovery method called Cell Regeneration on VP (CREG-VP) is proposed. One of the driving ideas of this proposal is to integrate the needed hardware in the network rather than in the customer premises equipment. To achieve the FEC scheme the authors propose a cell based matrix such as the one depicted in Fig 3.9 rather than an octet based one as in AAL1. An extra line and column of redundancy cells are added which are obtained by Xoring the data cells by line and column. The goal of a VP based error protection is to better protect against burst errors. Since the traffic flow is an aggregation of single Virtual Connections (VCs) the bursty cell loss model (Gilbert model) applies. The problem with this technique is that it begins to suffer from performance degradation for a relatively low link utilization of 75%. This approach does not specifically target real-time applications since the aggregated traffic is composed of all types of service.

Another technique targeting both data and real-time communications is the hybrid ARQ and FEC proposed in [95] and [96]. This hybrid technique consists of using FEC as the standard error corrections mechanism. Whenever, FEC is unable to recover the data, due to high losses in the network, the ARQ mechanism retransmits the missing data.

While all these contributions propose an AAL level FEC, an ATM Forum contribution [97] raised the question of the proper place of FEC. The authors argued that FEC is a physical layer issue. In fact if we consider the early applications of FEC the argument is valid because the majority of errors occurred at the physical layer. FEC was originally used in bursty error-prone networks which had important BERs. However in ATM networks the situation as already discussed is very different. In such environments the main source of errors is cell loss.

A set of contributions studied the feasibility of FEC based on RSE codes [98, 91]. The authors argued that the available technology allows to implement very high speed FEC. Several contributions claimed that a cell level FEC was a necessity for point-to-point and point-to-multipoint data transfer as well as for multicast communications [99, 100, 101, 102, 103].

An AAL-level FEC to improve both data and real-time communications is proposed in [99, 104].

Contributions against the utilization of FEC were also written. In particular, in [105] it is said that the greatest impact of FEC is when no retransmission is possible like in multipoint configurations. Also, considering that in general applications use SSCOP or TCP no other error correction is needed. This clearly reduces the discussion to data applications since such protocols rely on retransmission to correct errors. It is also stated that AALs have to be application independent. This is true however, FEC is not application dependent but targets a class of applications. The authors also argue that such FEC techniques are incompatible with EPD techniques which is true under the assumption that the applications that will make use of FEC run on top of AAL5. In addition, EPD leads to worse results for audiovisual applications. Another argument found in this contribution is that the cell loss models are bursty and under these conditions FEC is not effective. Our early discussion on cell loss model (see Sec. 3.2.1) shows that the opposite situation is more likely to occur and therefore refutes the argument. Finally, the authors also state that FEC has limited usefulness for low bit rate sources. This observation is again true assuming that block interleaving such as in AAL1 is applied. The ATM Forum's final decision was not to develop an AAL-FEC. One of the reasons for this, given the precedent overview, is that the scope of the proposed FEC mechanism was too large. Considering the different requirements and characteristics of audiovisual and data traffic, a single solution was difficult to find.

In fact, today there is no FEC-capable protocol able to handle simultaneously stream and frame oriented applications satisfactorily. The final development of this discussion has been the specification of a FEC-SSCS for AAL5 proposed in [100] and described in [104, 106] also proposed in [107].

One of the common points of almost all of these proposals is that they rely on the interleaving of a matrix of data. The problem with the approaches targeting MPEG-2 applications is that they do not take into account the nature and format of the data. As already seen in Sec. 3.3.1 not all the frames have the same importance and the impact of loss onto the decoded data is very different. In general the described proposals do not propose any adaptive mechanism but rather variations of AAL1's error correction mechanism. Also, the size of MPEG-2 Transport Packet sizes is relatively small. Bearing in mind the considerations of Sec. 3.2.1 the probability of observing multiple cell losses in a single packet is small and therefore the need of an interleaving to cope with bursty losses is arguable.

3.3.3 Transport Protocols for Multimedia

Still, not all the proposals for robust transmission consider only the use of FEC to improve robustness. Another approach consists in focusing on the transport protocol, who's main task is to provide a reliable end-to-end transmission regardless of the underlying network technology. Several multimedia transport protocols have already been proposed, in particular due to the interest of deploying multimedia services over the internet. Therefore, most of them targeted IP based networks. Consequently, the vast majority of the proposals do not take into account the characteristics of broadband networks [108].

The *eXpress Transfer Protocol*, XTP, was originally developed as a light-weight transport protocol for high speed processing. It is a connection-oriented protocol that covers layers 3 and 4 functionalities as shown in Fig. 3.10. It provides such features as a rate-based flow control, priorities, QoS parameters and selective error correction. It supports different types of service, including reliable transfer. The utilization of XTP on top of ATM is discussed in [109]. It stands out that some incompatibilities exist between both. In particular, the multicast support of XTP has been designed for broadcast networks. Every participant is assumed to receive all control messages issued by other participants. Another problem inherited from shared medium networks is the assumption of implicit connection setup. Once the setup packet has been transmitted, the connection is assumed to be usable, which is not the case in a connection-oriented network such as ATM. In fact, XTP mainly provides functionalities for datagram transmission and not for real-time applications which makes the mapping to ATM difficult.

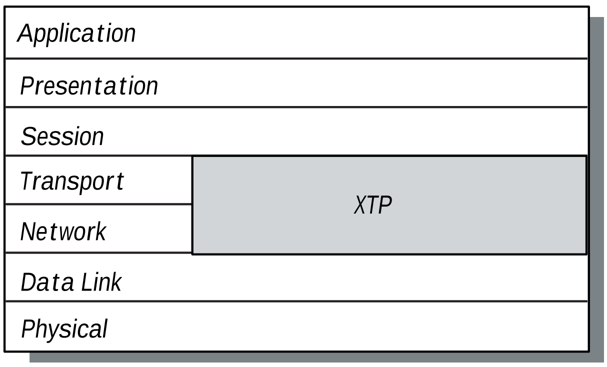


Figure 3.10: *The eXpress Transfer Protocol and the OSI layers.*

The work done within the *Heidelberg Transport System* (HeiTS) [110] framework aimed at developing a multimedia communication system for real-time delivery of video, audio and data in point-to-point and multicast environments.

The Internet Stream Protocol version 2 (ST-II) [111] was adopted as the network protocol for HeiTS. ST-II provides a connection-oriented service with a *sender-oriented* resource reservation mechanism. ST-II provides a large set of functions for the transmission of multimedia (See Sec. 3.3.4 for details).

HeiTP [112, 113] the Heidelberg Transport Protocol was designed to complete the functions provided by the network protocol. As such, HeiTP was designed to work on top of connection-oriented networks with multicast facilities and QoS guarantees. HeiTP provides a 1-to-1 simplex or full duplex connection as well as multidestination simplex connections. It does not perform any multiplexing. HeiTP allows to define per stream QoS and opens multiple network connections for multiple data streams. One advantage of this mechanism

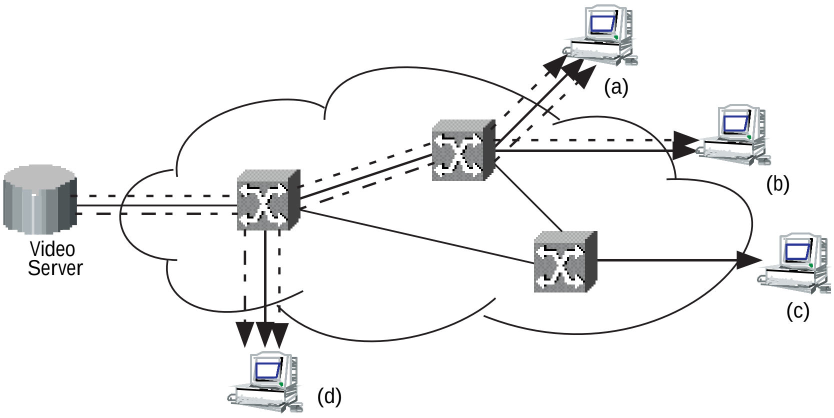


Figure 3.11: *Multiple video flows filtering principle.*

is to simplify the identification of packets at the destination since a one-to-one mapping between connections and flows is provided. To allow ST-II perform resource reservation, HeiTP passes QoS parameters in a flow specification format known as flowspec understandable by the network layer. This flowspec has also been included in IPv6. The congestion control is based on feedback information from the receiver. This feedback information is used by the transport layer to perform *media scaling* [110, 113].

Media scaling consists of dropping data when network congestion is detected. It is a rate control technique for lossy data such as multimedia streams which can accept some loss. By scaling down data flows, bit rates are decreased, therefore helping the network to recover from congestion. Media scaling is of two types; transparent and non-transparent. Transparent scaling works independently from upper layers and can drop parts of the data. The transport protocol performs the scaling on its own. Non-transparent scaling requires an interaction between the transport layer and upper layers. The transport protocol needs some knowledge on the data to be transported to perform an intelligent dropping.

The filtering methods depend on the nature of the data to be scaled. If the data is audio, then transparent scaling is difficult due to the high temporal coherence of the data. Therefore non-transparent scaling such as reducing the sampling rate can be applied. For video streams transparent scaling can easily be applied. Temporal scaling will reduce the number of frames transmitted by frame dropping. Spatial scaling will reduce the resolution in the video stream. Hierarchical coding has the advantage of giving different resolutions and thus downscaling can be done by dropping one of the highest resolution flows.

These two scaling techniques could also work in a continuous or discrete way. Continuous scaling consists of dropping the data from a single flow. The disadvantage of this technique is that it has an all-worst approach for multipoint connections. Conversely, discrete scaling relies on the utilization of multiple flows (e.g. data partitioning schemes for video). Discrete scaling will drop data from the lowest priority flow therefore reducing quality without disrupting the high priority flow. Fig. 3.11 illustrates this principle. Three flows are generated by the video server. Each has one level of priority. The base layer (solid line) guarantees a minimum quality while the enhancement layers (dotted lines) improve the image quality. In the example, two users, b and c cannot process all the incoming data. User b filters the third flow and receives the base layer plus an enhancement layer. User c is able to process only the base layer.

HeiTP provides 8 levels of priority. Bearing in mind that the impact that data loss has on video is extremely difficult to know, media scaling becomes very difficult to apply, not to mention the complex filtering mechanisms needed to perform the data selection. HeiTP provides also different reliability classes to handle different types of data. As a transport protocol, HeiTP provides detection of data corruption, duplication, misordering, loss and also *lateness*. Data loss and corruption is handled via retransmission. Time stamps are included in each packet to track the arrival times of the packets. Even if HeiTP provides interesting features for multimedia, it was clearly designed to work on top of ST-II which is an Internet protocol. If the design was supposed to work on ATM networks several features appear as redundant, such as the misordering error handling or much more important, the existence of ST-II agents to take care of the data filtering or the utilization of the flowspec to pass QoS specifications.

Recently, the IETF has developed a new transport protocol for multimedia, the Real-time Transport Protocol (RTP). This protocol aims at transporting audio and video over the internet and provides end-to-end mechanisms for timing recovery and error control. The relevance that RTP is getting with the fast development of multimedia services over the internet requires a separate description and is therefore presented in detail in Sec 3.5.4. It is foreseen as a transport protocol that would make use of ST-II as the underlying network protocol. The combination of both protocols would provide real-time multimedia transport on top of a connection-oriented network with resource reservation. Of course such service would be available when routers would support ST-II agents, filtering and resource reservation, features still not available today.

The described protocols have the disadvantage of being based on traditional data network concepts. The work described in [114] adopts a different approach. Based on the characteristics of broadband network such as a high bandwidth-delay product and low bit error rates they propose a new transport protocol called RAPID. This protocol addresses the transport of both multimedia and data traffic. It provides two different modes of transmission to adapt to both types of traffic requirements. The streaming mode, offers a low delay transmission for multimedia data. RAPID is able to provide per connection FEC on a packet basis which may generate a relatively large overhead. The difference between RAPID and other transport protocols is that it does not provide retransmission error correction albeit it provides a rate control which leads to low delay transmissions but to a much more unreliable service especially for data transfer applications.

Even if there is no exact mapping between the OSI layers and ATM, the AAL provides part of what are considered as transport functions, which is the reason why AAL proposals are included in this section.

A specification of a new AAL for video is presented in [115]. The driving idea in the paper is that current AALs are not adapted to video transmission. AAL1 generates delay, overhead and provides a dejittering function sometimes considered as superfluous at this layer, in addition to the fact that it provides CBR services only. On the other hand, AAL5 provides a frame-based protocol which potentially increases the cell loss to the frame level. Since AAL5 has no means to recover lost cells, packets are discarded thus increasing the loss ratio experienced by the receiver. The author proposes a generic video adaptation in the sense that only a basic cell loss detection algorithm, which he identifies as the main problem in current AALs, will be performed. Resynchronization, jitter-removal and bit errors are considered as application-only functions. To provide this functionality, two SAR-PDU formats are proposed: the first uses a single byte header. It includes a sequence number and a CRC. The second uses two bytes for the header. It includes a sequence number (SC) an octet counter (OC) a segment length (SL) field and a CRC. The OC field transmits the number of previously transmitted octets and the segment length is used for partially filled cells. The combination of OC and SC allow to detect cell losses. Partially filled cell functions

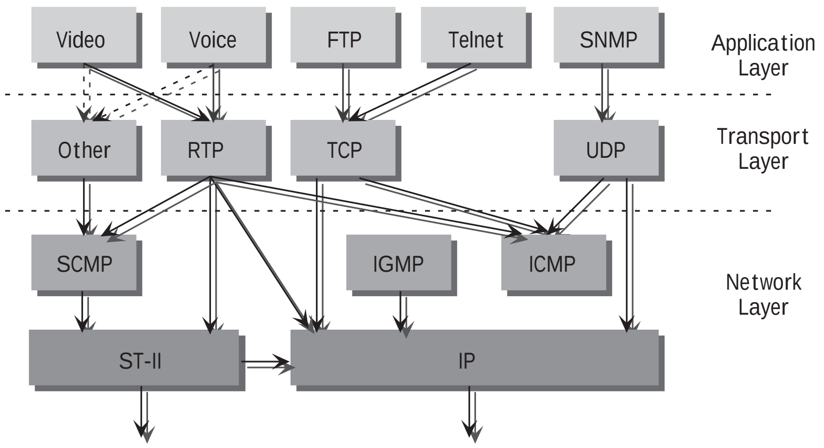


Figure 3.12: *Internet protocols relationship.*

are foreseen to avoid excessive packetization delays in VBR video applications as well as for the transmission of unpacketized video streams other than MPEG-2. To allow the receiver to detect the different cell formats, the ATM User-to-User bit of the cell header is used. The advantage of this proposal is to provide a simple but efficient cell loss detection mechanism. However, it generates some extra overhead due to the headers and it does not provide any cell loss recovery mechanism.

3.3.4 Network Protocols for Multimedia

Network protocol layers do not provide for reliable transfer. However, some network layer mechanisms such as resource reservation and multicast support may help to increase the quality of service offered to the applications.

The most important network protocol layer developed for multimedia is the Internet Stream Protocol. Recently, the IETF has developed ST-II+[116], a revision of the original proposal to improve interoperability and to eliminate original features considered as useless today.

ST-II is a *connection-oriented* internetworking network protocol. The main development goals were the support of efficient data steam delivery to single or multiple destinations in applications requiring guaranteed throughput and low delay. ST-II is considered as a multimedia protocol. ST-II is able to reserve bandwidth across network routes if, obviously, the routers implement ST-II. Given that multimedia applications are foreseen as the main users of multicast, one of the key features of ST-II is its ability to support multicast communications.

ST-II has been designed to coexist with IP on each node. Since multimedia combines audiovisual information as well as data both network layers would be used by the applications. However, ST-II will not make use of the same transport protocol as IP as depicted in Fig. 3.12. Both ST-II and IP use the same addressing scheme and have the same packet structure. The main difference is in the version number which is number 5 for ST-II and number 4 for IP. Note that this explains why IPng is now know as IP version 6.

ST-II is based on the concept of streams. A stream is composed of:

- a set of paths between the origin (source) and the targets (destinations)
- a set of resources allocated through the nodes by *ST agents*

- state information maintained at the nodes.

ST-II streams are directed trees between the origin or root and the destinations or leaves. As such, streams are called multidestination simplex, since the data flows only downstream. This is not the case for the control information that also flows upstream.

To allocate resources in the nodes, QoS parameters are negotiated at connection setup. These parameters form a *flow specification* or *flowspec*, concept also available in IPv6. A flowspec which may contain delay, average and maximum throughput for example is transmitted to all nodes and leaves. Each node contains an ST agent which is in charge of allocating resources and also maintaining state information. The receivers are able to accept or refuse the connection based on the flowspec which they could not modify, therefore making ST-II a *sender-oriented* protocol. Even if ST-II sets up the paths between the origin and the targets it is not a routing protocol. It accesses routing information via the ST agents to perform the resources reservation and data path selection.

The last main feature of ST-II to support multicast is the concept of *filters*. Considering that in multipoint configurations receivers will not have the same capabilities, the approach while designing ST-II to support heterogeneous terminals has been to use filtering mechanisms to adapt the data flows to the terminal's capabilities. Unlike the media scaling mechanisms described in Sec 3.3.3 developed for transport protocols, the filtering mechanisms proposed at the network layer target packets. Routers, implement protocol stacks up to the network layer. They could in any case know the nature of the data they are processing and therefore media scaling could not be performed on this basis. ST-II proposes a *data priority* mechanism to perform packet dropping. The packets are *tagged* according to a level of priority. In case of congestion, low priority packets will be dropped first to guarantee a minimum QoS. Of course, this type of filtering is achievable if the data is encoded according to a hierarchical scheme such as the scalability profiles proposed for MPEG-2.

Albeit ST-II provides interesting features for multimedia communications it has seldom be deployed. In addition, IPv6 combined with RSVP offer a similar but much more flexible framework that will certainly replace ST-II.

Goyal *et al.* propose another network protocol layer for video transport. This layer guarantees zero loss for VBR video applications. To achieve that, the layer uses a hop-by-hop credit-based flow control to reserve buffer space for a communication channel in all the switches along the path. To ensure low latency and jitter, the proposed protocol dynamically allocates the bandwidth based on a real-time estimation of the bandwidth requirements. However, due to the nature of the proposed mechanism, no bounds on the delay are guaranteed. This proposal reflects the trend of using ABR services for video transmission. However since no guarantee on the delay is given this approach precludes the use of ABR, at least for interactive communications. The prediction algorithms proposed are specific to video which results in a video only ABR mechanism.

Finally, a *channel adaptation* layer is proposed in [117, 118]. This new protocol layer aims at providing an integrated approach towards the transmission of MPEG-2 over ATM and therefore cannot be fitted in any specific protocol layer. It discusses the options for multiplexing MPEG-2 data over ATM. The authors propose a *user multiplex* channel adapter and a *channel multiplex* adapter. The former is a direct mapping of a single bit stream into ATM cells. Both a loose and a tight packing schemes are proposed. The latter proposes the mapping of different streams into separate VCs. The allocation is performed on a PID basis. One of the advantages of this technique is that it allows to route also PSI or Program Clock Reference data into separate VCs therefore breaking the bit stream into raw data and temporal information. However, separating the temporal information from the video data requires the receiver to do an extra synchronization between both flows. The major drawback of this proposal is the high level of integration. The channel adaptation layer includes functions of the ATM, AAL and higher layers making it usable for MPEG-2 applications over ATM exclusively.

Another integrated approach is presented in [119]. The authors propose the encapsulation of MPEG slices directly into CS-SDUs. The CS header information is generated with a specific knowledge of the MPEG data structure. The same holds for the SAR sublayer which makes the AAL totally MPEG specific. Their approach is based on an early study of video data packing in ATM cells [75]. Both tight and loose packing schemes of slices are proposed. An interesting idea presented is the interleaving of odd and even slices. This technique should avoid the loss of subsequent slices, thereby increasing the efficiency of simple spatial interpolation. The drawback of this proposal is that the protocol layer which includes transport and AAL functions is totally MPEG specific and the slice interleaving scheme is also not conforming to the MPEG standard finally making the proposal totally proprietary.

3.3.5 Error Concealment

The third technique to reduce the data loss impact is the error concealment. Even if error concealment is not directly related to the transmission itself some networking mechanisms can improve the efficiency of error concealment. We present here a short overview of the domain. A complete review of the topic can be found in [120, 121].

The main purpose of error concealment is to mask the errors into the decoded image. Error concealment is based on the property of masking of the human visual system. Errors in a bitstream and in particular data loss, cause visually perceptible errors which propagate in space and time. The loss of an ATM cell containing a little portion of a single picture can be extended into much larger areas. This is due to the behavior of the current family of decodesr when data loss is detected. Normally, a decoder will temporally stop decoding up to the next resynchronization point (startcode) which in general is a slice header. Since the lower level synchronization points are at slice headers the rest of the slice affected by the error will be discarded. This is visually seen as a black strip in the image. Two techniques are applied to limit the extension of areas damaged by errors. The first is based on interpolation. The idea is to use the surrounding information to interpolate the missing data. The second is the early resynchronization. The technique consists in forcing the decoder to resynchronize at macroblock level without waiting for the next startcode.

The interpolation techniques are of two types; spatial and temporal interpolation. The spatial concealment predicts the lost macroblocks by linear interpolation of pixels from the nearest pixels of neighbor macroblocks correctly decoded. Since generally the available data is above and below, the interpolation is vertical(see Fig. 3.13). This technique works with all types of pictures. Another spatial technique for I frames only is done in the frequential domain, where DCT coefficients from neighbor blocks are used for interpolation. The uses the temporal correlation of images. The principle is to substitute the damaged areas of an image by the contents of the corresponding blocks in the nearest temporally decoded picture as depicted in Fig. 3.13. This technique works also on any kind of picture.

The early resynchronization technique uses a different mechanism. It limits the spatial propagation of errors by decoding some semantic information that is normally discarded from the damaged MPEG-2 video streams. Basically, it tries to use all the correctly received data. In other words, it helps the decoder to resynchronize quickly (at least before the following header). This method is based on the identification of allowed codewords as

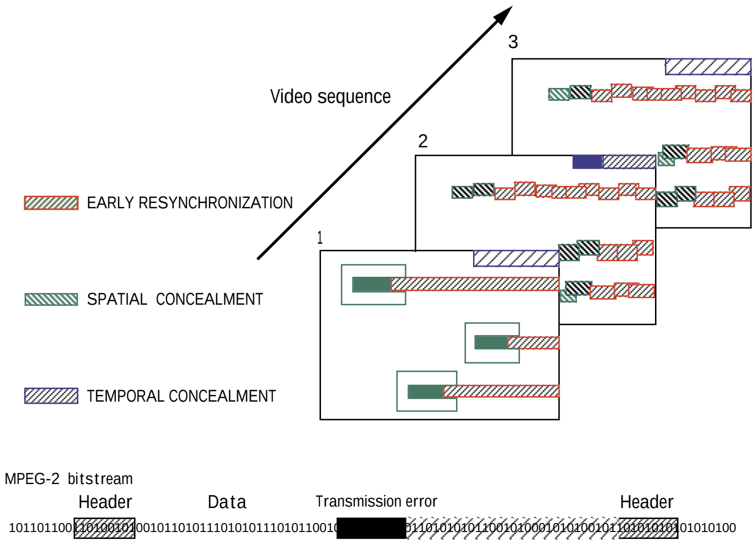


Figure 3.13: *Error concealment techniques.*

proposed in [122] and, unfortunately, works only with intra-coded frames. This technique is based on the assumption that the first octet received after an error occurrence corresponds to the beginning of a new macroblock. The decoder will try to decode all the information on this basis. If an invalid codeword is detected, the decoder discards the decoded information and retries resynchronization for the next octet. This process is repeated until syntactically correct macroblocks are decoded as shown in Fig. 3.14.

Figure 3.14: *Early resynchronization example in a basketball frame.*

The fundamental difference between the interpolation techniques and the early resynchronization is that the latter tries to use all the available information. In order to maximize the available information at the decoder, it may be interesting that the lower layers deliver all the packets even if they are incomplete rather than discarding them prior to decoding.

All these techniques are, by far, not perfect. Data loss may still involve very annoying perturbation in the decoded video, especially when frames are lost.

3.4 Multipoint Communications Requirements

Multimedia applications over B-ISDN will extensively use the multipoint capabilities. Entertainment services including near-VoD are based on multicast and broadcast functionalities (see Fig. 3.15). Its possibilities will also be applied to enterprise services for collaborative work and video conferencing. The education community will also make use of applications such as teleteaching [123], interactive distance learning and telemedicine. However, several issues are still to be solved to have a reliable multipoint, multicapabilities service over ATM. Three main issues are under research today, namely:

- how to efficiently establish multipoint-to-multipoint connections
- how to efficiently support multiple QoS requirements
- how to guarantee network QoS for such connections.

The first two items are to be covered by the control plane that handles the connection establishment and the network resources allocation. The third issue is to be handled by the user plane that takes care of the data transport and delivery. We describe the three issues in the remainder of this section. However this work focuses only on the user plane aspect of the problem and proposes a solution for data transmission only.

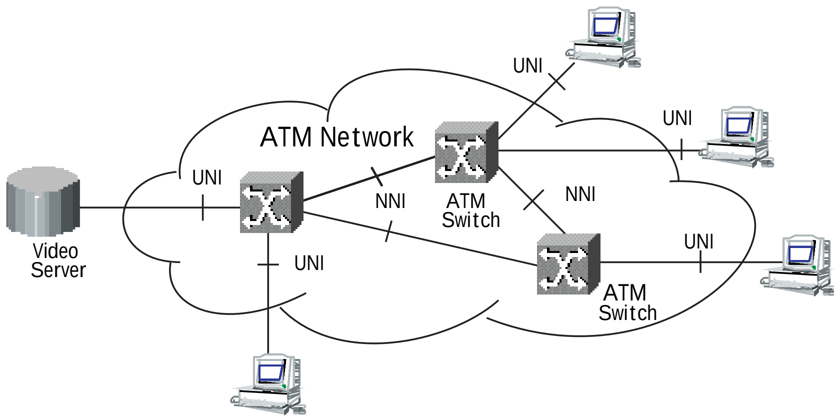


Figure 3.15: *Multicast network example.*

3.4.1 Multicast in ATM Networks

The problem of multicasting multimedia services over ATM is inherent to ATM philosophy [124] and to real-time audiovisual services nature. In shared media networks such as Ethernets and token rings all the workstations are attached to the same LAN segment and so process all the packets in that segment allowing to send and receive data to and from multiple end users simultaneously. There is no analogous solution in ATM. AAL5 does not provide the mechanisms to receive interleaved packets coming from different sources. AAL3/4 provides a Message Identifier (MID) field of 10 bits (see Sec. 2.2.3.3) that precludes the interleaving problem but still has other drawbacks; there is no mechanism to ensure the uniqueness of the MID value within a multicast group and the reduced size of the MID field will severely limit the number of possible participants. This is the main reason why multicast on ATM is point-to-multipoint only. In addition, AAL3/4 has considerable overhead and is relatively complex to implement.

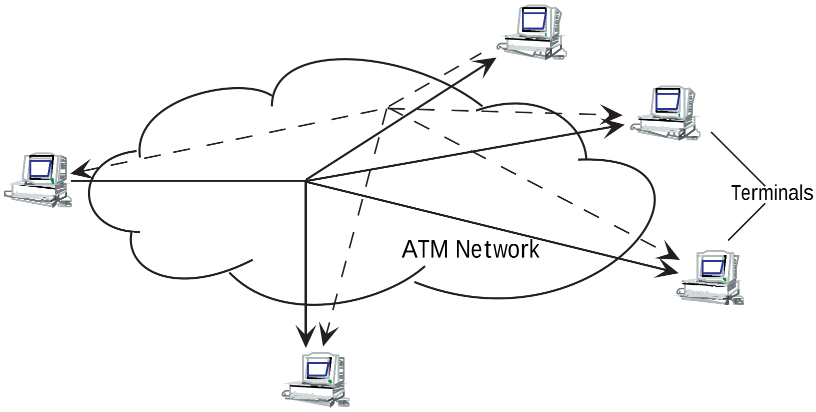


Figure 3.16: *N-Point-to-Multipoint configuration example.*

For these reasons, the configurations adopted nowadays are such as the one shown in Fig. 3.16 where a single multipoint-to-multipoint communication is replaced by N point-to-multipoint connections. From a technical point of view, this solution is relatively easy to implement but depending on the application it can lead to user problems. The management of a multiparty conference based on the configuration depicted in Fig. 3.16 can become difficult if leaf initiated joins [42] are permitted. Clearly, this situation is suboptimal from the network resource allocation point of view. Fig. 3.17 depicts a three party configuration. If we assume an audio conference, having three point-to-multipoint connections, this leads to the situation of Fig. 3.17 (a). Each party has an outgoing connection and two incoming connections. If we assume that each connection uses a bandwidth b, $3b$ is allocated per participant. If we take into account that in such conferencing not all the users speak simultaneously, there will always be $2b$ unused. If a multipoint-to-multipoint connection would have been established, it could have been done with an allocation of b, leading to a considerable saving of bandwidth as depicted in Fig. 3.17 (b).

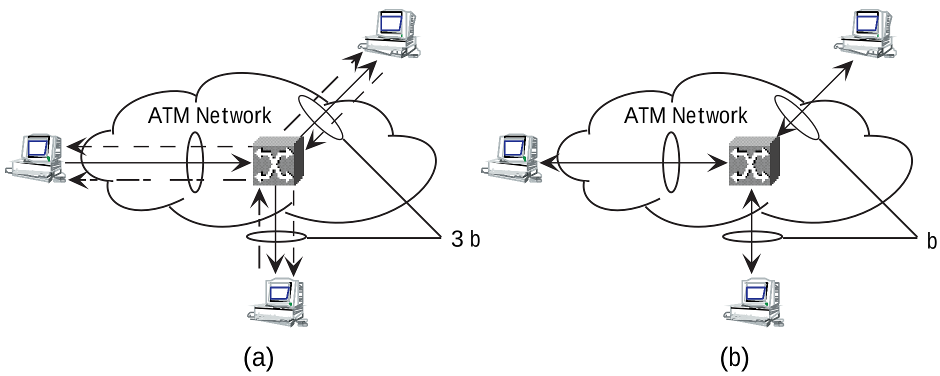


Figure 3.17: *N-Point-to-Multipoint vs. Multipoint-to-Multipoint.*

3.4.2 Supporting Multiple QoS in Heterogeneous Environments

ATM is a connection-oriented technology, and as such it has the ability of providing per-connection QoS. This feature becomes difficult to manage in heterogeneous environments such as they can be found in multipoint connections because ATM point-to-multipoint connections are able to deliver a *single QoS* for all the receivers. This aspect of ATM can become a problem in configurations such as the one depicted in Fig. 3.18. In this video broadcast configuration, the video server is able to deliver three levels of quality; a base layer and two enhancement layers. Each of the receivers has different capabilities and therefore different requirements in terms of QoS. The first terminal is able to decode the three layers to obtain maximum audio and video quality. The second end-user can decode one of the enhancement layers and the third receiver can decode only the base layer to have a standard quality.

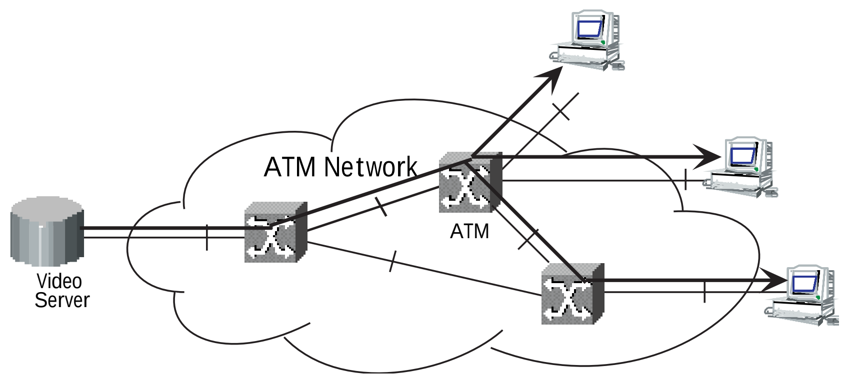


Figure 3.18: *Video broadcast configuration.*

The user with the low-end Set Top Terminal (STT) will not want to pay for extra bandwidth that cannot be exploited and the user with the high-end STT will not want to receive a standard TV quality if he is paying to receive a full quality program. The simple solution would be to establish three point-to-point connections each of which will have its own QoS. While this solution is very easy to implement today and will certainly keep the users satisfied, it is extremely inefficient in terms of network resource allocation: bandwidth is used three times to send the same base layer to the users. If exploited in a more efficient way, the network provider will be able to allocate part of the bandwidth to other users and services. On the other side, the only way of delivering the data to all three users is to rely on the lowest quality receiver.

The research community has already addressed the problem of efficient resource allocation in multipoint communications with heterogeneous terminals, principally the Internet community. The main solutions envisaged are based on the deployment of resource reservation protocols [113] combined with adapted Transport layer protocols. There are two main types of reservation protocols; sender-oriented [125] such as ST-II (see Sec. 3.3.4) which in fact is a network protocol with resource reservation mechanisms and receiver-oriented [2] such as the *Resource Reservation Protocol* or RSVP.

The sender-oriented protocols are based on a sender QoS specification or flowspec transmitted to the intermediate nodes and the receivers. The source is in control of the connections. It can change the flowspec if necessary, add and remove receivers and so on. The agents in the nodes and the receivers on the multicast tree are only allowed to accept or refuse the connections as well as the modifications of the flowspec. Clearly, such resource reservation philosophy is oriented towards specific services in which every receiver has the same capabilities and is able to receive all the data. As an example, today's video distribution systems have such configuration. One of the problems of sender-oriented protocols is that the source may eventually become a bottleneck if too many receivers get connected.

Receiver-oriented protocols such as RSVP are also based on a flowspec. The difference is that the QoS requirements are sent by the receivers to the source. This does not change at all how the source behaves, in fact, the source delivers a single QoS and it is up to the intermediate nodes via agents to do the filtering (see Sec 3.3.3) and adapt the data according to the QoS requested. The advantage of having the intermediate nodes participating to the QoS adaptation is that they can adapt dynamically to local network conditions and also avoid the feedback implosion problem by intercepting the requests sent by the receivers. This resource reservation scheme is adequate for configurations whether the receivers have different QoS requests.

While media scaling based on agents can be relatively easy to implement in IP based networks, it does not seem to map well to ATM. Indeed, ATM switches implement the physical and the ATM layers. Their task is to route cells to the output port associated to the VPI/VCI values in the ATM cell header. But to provide some filtering, the switch has to know what kind of data it is receiving and this implies implementing at least the AAL to reassemble packets and be able to gather some information to decide whether the data has to be filtered or forwarded. This will make ATM switches behave much like routers. An implementation of such kind of solution could be done in the same way it has been done to transmit connectionless data over ATM; by coupling a connectionless server (CLS) to the switch [38, 39]. A CLS reassembles the data (e.g. IP packets) and maps the packet IP address to a VPI/VCI address by consulting a table. This solution, although cumbersome, is viable for file transfer applications that do not have timing constraints but cannot be envisaged for real-time applications. It will moreover generalize the solution of having external specialized devices for particular applications. Still, there is an alternative to improve speed in CLS: the streaming mode. This mode does not reassemble the full packet. It checks the IP address on the first cell of a packet and then routes all the cells belonging to the same packet. This solution is not applicable to multimedia applications because it implies that the device which does the routing or the filtering will had to know which information is it looking for. This leads to a specialized device that will depend on the codec used by the applications. Moreover, the information to be filtered is not necessarily contained in a single packet (e.g. a frame).

Another solution already applied is the Early Packet Discard (EPD) mechanism. EPD, allows to selectively discard cells from the same packet in case of congestion. Again, this solution fits data transfer applications because the transport protocol will notice that a packet is missing and will ask for retransmission. As far as video is concerned, discarding a packet will amplify the loss seen by the application and the impact of this mechanism will depend on the size of a packet (e.g. a Transport stream, a complete image, etc...)

A different approach is described in [126]. The authors propose a control protocol for shared ATM multicast trees which is able to support multipoint-to-multipoint communications. SMART is based on the ABT concept where series of cells are delineated by RM cells. The utilization of a token-passing algorithm to control transmission over the multicast tree allows to achieve multipoint communications with an unbounded number of users using only two virtual connections. The advantages of SMART are the following:

- requires only two multipoint-to-multipoint connections
- does not require any external device such as a broadcast server
- does not require n point-to-multipoint connections for n end-users

- does not require a MID field so any AAL can be used

- guarantees that no interleaving of data occurs.

Since SMART is a control protocol for accessing shared multicast trees, it does not deal with data transmission and therefore cell switching is performed at the ATM level. SMART works with guaranteed QoS connections such as CBR and VBR through a *static reservation* mechanism as well as with dynamic ABR-like traffic through *dynamic reservation*.

However, the proposed protocol still has a drawback, which is common to most of the existing solutions, it requires the network elements to implement the protocol. The token-passing algorithm requires an arbitration that is performed by the switches.

The deployment of such solution therefore requires an endorsement from the specification bodies to make it a standard that will allow for its deployment.

In fact, the situation today does not necessarily need such kind of approach. In the coming years, networks will be heterogeneous mixing ATM and IP principally leading to an *IP overlay network* paradigm. This calls today for solutions able to interwork. If RSVP becomes the internet standard for resource reservation, filters will not be mapped to ATM. The filtering functions will be done within the IP network, the ATM being seen mainly as a high speed backbone for the internet. What needs to be defined to make such networks work would be the mapping of FlowSpecs and resource reservations to ATM. In a sense, ST-II is better suited to interwork with ATM because of its connection-oriented nature, easy to map to ATM VPI/VCIs and also due to its multidestination simplex connection which easily maps to ATM point-to-multipoint connections. SMART also appears to be a good solution if QoS mapping is done.

3.4.3 Multiple QoS Guarantees in Heterogeneous Environments

Resource allocation tries to solve the problem of providing different QoS to heterogeneous terminals. However, it does not solve the problem of maintaining the network QoS required by multipoint connections (see Sec. 3.3). In general, when cell losses or too long delays occur, image and sound quality degrades. Two approaches are possible to tackle this problem: the first consists of applying mechanisms to detect and recover from cell losses assuming that the application cannot handle all the errors by itself. The second is based on the opposite reasoning, that is to rely on the application to cope with all the errors. This applies only for audiovisual applications that do not use reliable transport protocols. The problem with the second approach is that the application is not a part of the network; it uses it and as such is expecting the QoS required. If the network wants to really guarantee a given QoS even in case of congestion or loss it has to apply some mechanism to compensate. The situation in multipoint connections is that the network has to guarantee simultaneously timing constraints and multiple QoS for multiple end-users. In such a configuration the error recovery mechanisms have to be independent of the source, especially in large multicast groups, to avoid the feedback implosion paradigm which in turn generates too large delays when applied. If we consider interactive multimedia services the only possible solution is the utilization of FEC techniques. Since the recovery is done at the receiver only, two advantages appear: there is neither acknowledgment mechanism nor retransmission delay in case of error thus minimizing the end-to-end delay even in case of congestion. This alleviates the source from any extra load. In return, FEC techniques add an extra overhead which if not carefully applied may contribute to overload the network in case of congestion.

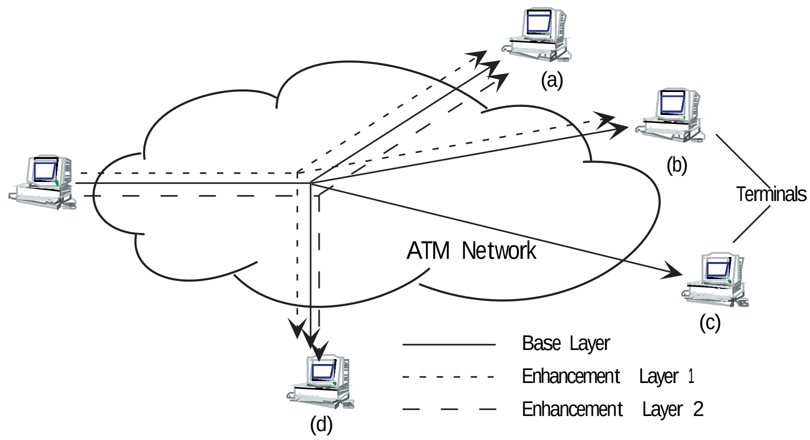


Figure 3.19: *Multiple multicast trees connection.*

3.4.4 Integrating Multimedia and Multicast Configurations

The problem of multipoint services is a complex one. A multipoint service has to be designed to cope with heterogeneous terminals, different QoS requirements and also different network conditions within a multicast tree. However, it is possible to achieve a reliable multicast for hereogeneous terminals by integrating the different aspects of multimedia applications and ATM networks as follows:

- layered coding as discussed in Sec. 3.3.1 allows to separate the multimedia information in a base layer and one or more enhancement layers
- a multicast connection in ATM is able to deliver a single QoS
- a multicast connection in ATM uses one ATC.

It is therefore possible to combine all these three elements to create different QoS multicast trees as shown in Fig. 3.19. The base tree delivers the base layer over a guaranteed ATM connection to all the users. Enhancement multicast trees could be established that will bring the enhancement layers to the terminals with enough capabilities. In the configuration depicted in Fig. 3.19 three layers are delivered by the source. Three multicast trees are establised. Terminals a and d receive the three layers, terminal b two and terminal c receives only the base layer. Finally, it could be envisaged to reduce the connection price by establishing the enhancement multicast trees over non-guaranteed or partially-guaranteed connections, such as UBR or ABR.

From the point of view of the data transmission, for sender-oriented as well as for receiver-oriented configurations, the use of FEC seems to have more advantages than drawbacks. Error correction does not need any feedback, thus reducing the delay, the load at the sender and at the upstream traffic. It does not require any feedback and therefore works over any multicast configuration. Last but not least, it does not rely on any network element and may be used when local network conditions require it.

3.5 The Standards for Multimedia

The standardization of an integrated services network started in the early 1980s with the development of ISDN technology. The target was to integrate narrowband services, especially telephony and data services. Later video conferencing services started to appear with the development of low bit rate video compression algorithms. The development and standardization of the Synchronous Digital Hierarchy (SDH) paved the way to optical communications and enlarged the objectives of ISDN by the inclusion of broadband services. ATM was chosen as the core technology for the B-ISDN and standardized by ITU-T. Simultaneously, video compression algorithms such as H.261 and MPEG were standardized by ITU and also other standardization bodies such as the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC). In the early 1990s, new standard groups such as the ATM Forum, the Internet Engineering Task Force (IETF) and the Digital Audio Visual Council (DAVIC) have fostered multimedia services by taking active part in the standardization of ATM technology, audiovisual communications and the internet.

Albeit very advanced, the standardization of so many different fields is far from being completed. The next sections summarize the current status of the different standard bodies related to the development of multimedia services and applications.

3.5.1 ITU-T

The International Telecommunications Union (ITU) edited in 1993 the biggest part of the I series of recommendations covering the specification of ATM. Among the four AALs defined in I.362, AAL1, AAL3/4 and AAL5 were already specified albeit some not completely. AAL2 which is still not yet specified today, was only defined probably due to the lack of class B applications.

With the development of real-time CBR services, particularly driven by video applications, ITU-T advocated the utilization of AAL1. SG13 Q.6, concentrated its efforts towards a final release of AAL1 in support of class A services. In particular, the delay created by the standard FEC interleaver was identified as a potential problem. In November 1994 a draft revision of I.363 named I.363.1 [37] which covered AAL1 and 2 only was issued. It included specific functions to support CBR video transport such as the transmission of AAL user information to the upper layers for error concealment and a short, low delay, interleaver. However, given the developments in the ATM Forum for VoD services and considering that the presence of AAL5 is required in all terminal equipment for signaling purposes prompted the possibility of using AAL5 for the transport of CBR MPEG-2 VoD to keep customer equipment cost low. SG13 Q.6 agreed in July 1995 to describe AAL5 as an alternative solution in addition to AAL1 for the transmission of constant packet rate (CPR) MPEG-2 coded VoD services. As a consequence of this choice, the last revision of I.363.5 [127] issued in May 1996 includes the possibility of passing corrupted PDUs to the upper layers. This extension however does not change very much the situation in case of loss. The document specifies that packets with a CRC error check but with a correct length indication are to be passed to upper layers which excludes incomplete AAL5-PDUs. It is worth to note that ITU-T decided to split I.363 into separate components each of which covers one AAL.

Since multimedia applications generally involve the transmission of synchronized distinct flows, the need for multiplexing and synchronization functions was identified. Recommendation H.222.1 [128] specifies a layer called *Network Adaptation* (NAL) which deals with multiplexing and synchronization of multiple multimedia signals (audio, video and data). The current version of the recommendation clearly states that the scope is restricted to

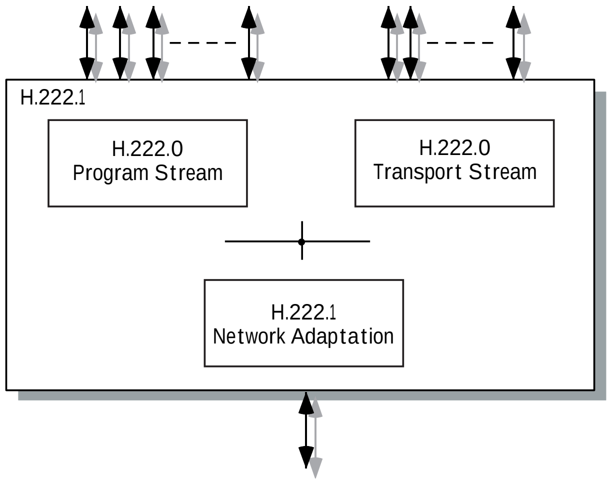


Figure 3.20: *H.222.1 overview.*

CBR over AAL1 and AAL5. The inclusion of VBR support is foreseen for further releases of the specification. The current version of this recommendation provides among others the following functions:

- multiplexing: multiple flows carrying data from one and only one media source can be multiplexed. This function precludes multiplexing at the MPEG-2 system layer because such a multiplexed flow will not be conforming to the recommendation
- timebase recovery: this function is provided to single media flows which have a single (common) timebase
- error reporting: it passes the AAL error report information
- priority: two types of priorities may be applied to NAL-PDUs.

Another reason to the existence of this layer is the lack of timing and error recovery functions of AAL5. This layer should provide the missing functions of AAL5 for the transmission of video. The recommendation says that this layer may provide dejittering if the CDV is beyond the limits accepted by the MPEG-2 system layer (H.222.0), the function being left as an implementation choice. Also, in early stages of the specification, the utilization of FEC at this layer was proposed. It was however abandoned because of jitter problems and also because such a choice was inconsistent with the AAL-CS model approach.

ITU-T Study Group 15 (SG-15 video experts group) developed recommendation H.310 that covers the technical requirements for the Broadband Audiovisual Communications Systems and Terminals.

It defines both uni-directional and bi-directional terminals. Their classification into different types is based on audiovisual and ATM Adaptation Layer capabilities. Two classes of uni-directional terminals have to be distinguished: Receive-Only-Terminals (ROT) and Send-Only-Terminals (SOT). Bi-directional terminals are referred to as Receive-And-Send-Terminal (RAST) types. All RAST terminals have the following interoperability principles:

1. Interworking between H.310 and H.320 terminals is mandatory.
2. Interworking among the different H.310 terminals is also mandatory.

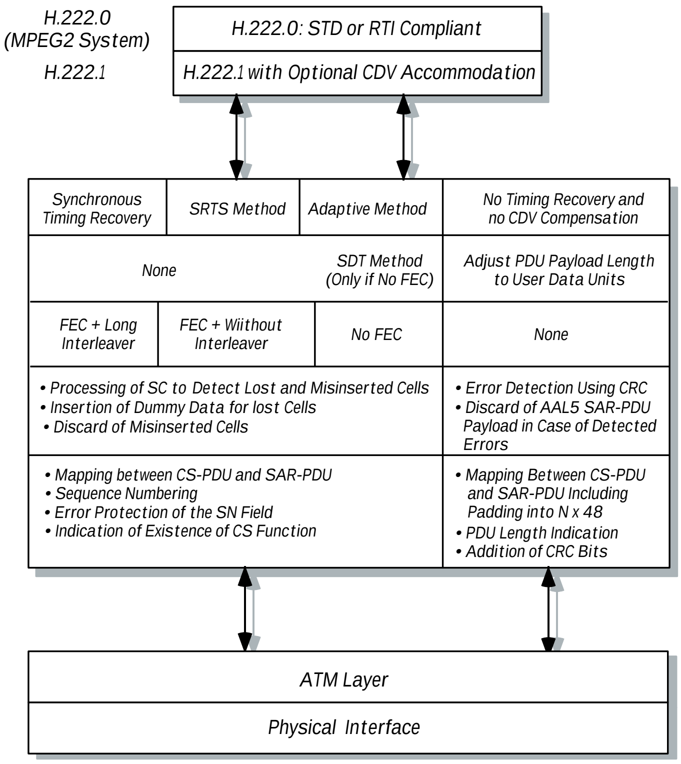


Figure 3.21: *Overview of the AAL1 and AAL5 functions for H.222.1.*

The specification of these terminal types is intended for the support of the following applications:

- conversational services (e.g. videoconferencing and videotelephony)
- distribution services with individual presentation by the recipient (e.g. VoD)
- distribution services without individual presentation by the recipient (e.g. TV Broadcast).

Both unidirectional and bi-directional terminals are defined according to the AAL capabilities supported (see Table 3.3). Each of the terminals have to support the so called H.310 *native communication mode*. This mode consists of H.222.1, ISO/IEC 11172-3 Layer II, H.262 and H.245 as the audio, video and control protocols. In addition, the RAST protocols have to support H.320/H.321 interoperation mode.

	AAL1	AAL5	AAL1&5
ROT	ROT-1	ROT-5	ROT1&5
SOT	SOT-1	SOT-5	SOT1&5
RAST	RAST-1	RAST-5	RAST-1&5

Table 3.3: *Definition of H.310 terminal type.*

The H.310 terminals have to implement H.262 Main Profile at Main Level video capabilities and also be able to handle hierarchical coding. Since RAST interoperability with

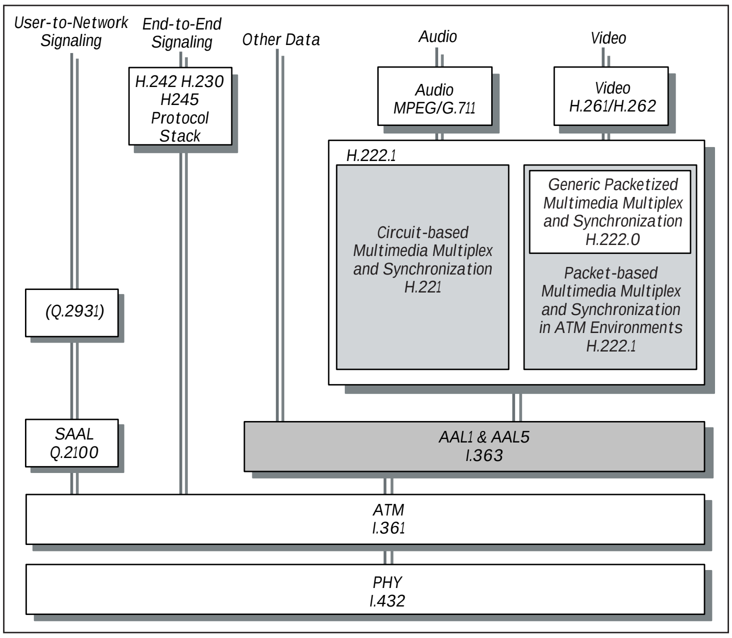


Figure 3.22: *Broadband audiovisual communication system and terminal configuration.*

narrowband terminals is mandatory they also have to support H.261 and H.263 video coding standards. The same situation holds for audio. MPEG audio is mandatory for all terminals, but RAST have also to implement G.711.

The utilization of H.222.1 naturally restricts the classes of service supported by these terminals to class A only albeit both AAL1 and AAL5 may be used.

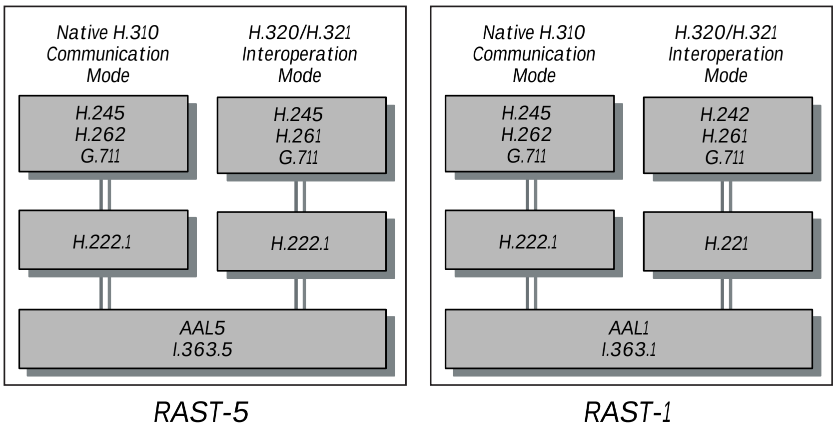


Figure 3.23: *RAST protocol stacks for native H.310 communication mode.*

3.5.2 The ATM Forum

In 1993 the Service Aspects and Applications (SAA) group of the ATM Forum started to define phase 1 of the implementation agreement for the carriage of audio, video and data over ATM in support of *Audiovisual Multimedia Systems* (AMS).

The first set of contributions defined the scope of the document in terms of services to include Multimedia Desktop , Video Conferencing , Interactive Distance Learning and Video on Demand [129, 130, 131, 132]. The final scope of phase 1 was however narrowed to MPEG-2 CBR encoded Video on Demand services. The Technical issues to solve were:

- how MPEG-2 TS packets are mapped to ATM cells
- how timing recovery is performed by the applications
- how error correction is performed
- how to reduce the impact of cell and packet loss.

All these issues are strongly related to the choice of the AAL and also to the type of traffic, CBR or VBR, to cope with. Some contributions proposed the development of AAL2 to handle VBR multimedia services [90, 133] and even an AAL6 specific to MPEG-2 services was proposed [134]. Several contributions, suggested the use of available AALs without precluding further developments towards AAL2 and an AudioVisual Service Specific Convergence Sublayer (AVSSCS) [135, 136, 89]. The first baseline text [137] issued in 1994 proposed an architecture based on the draft recommendation H.32x (see Fig. 3.24), which will later become H.310. This Framework did not preclude the utilization of any AAL. However, to have an implementation agreement ready within a reasonable timeframe the only issue was to use the available AALs.

The first issue studied by the working group concerned the problem of CDV accumulation and jitter removal. Multimedia applications require a short, bounded delay, as close as possible to a constant value. ATM networks introduce jitter basically due to queueing delays in the switch elements. CDV is a complex function that depends on several parameters (e.g. the number of switches in the path, the load and traffic profile, etc...). There is no simple and accurate model to describe the CDV.

The problem introduced by the CDV for isochronous services has already been tackled. One of the services provided by AAL1 is the circuit emulation for telephony. It has to emulate circuit switched networks so as to achieve a constant delay. AAL1 provides mechanisms for timing recovery: the Synchronous Residual Time Stamp (SRTS) [31]. Concerning VBR services, there is actually no satisfactory solution similar to SRTS. Contribution [138] proposes an adaptation of SRTS to VBR but the mechanism has not been tested yet and needs further mathematical developments for validation. Therefore, the specification restricted its focus to CBR services to reduce the complexity of the implementation agreement. The major problem with VBR traffic being that no efficient timing recovery algorithm was available. Some contributions proposed a modification of the adaptive mechanism of AAL1 but they were not accepted. As a consequence, AAL2 was not considered anymore. Two options were therefore available AAL1 and AAL5. As already discussed in Sec. 2.2.3.1, AAL1 provides error detection, optional error correction, and also time recovery. Moreover, it allows to pack an MPEG2 transport stream packet into 4 ATM cells without any padding. On the other hand, AAL5 provides error detection via CRC-32 parity check code, no error correction and no timing recovery. The mapping of a TS packet is inefficient since it needs 5 ATM cells with 44 bytes of padding. Another alternative is to pack 2 TS packets into 8 ATM cells

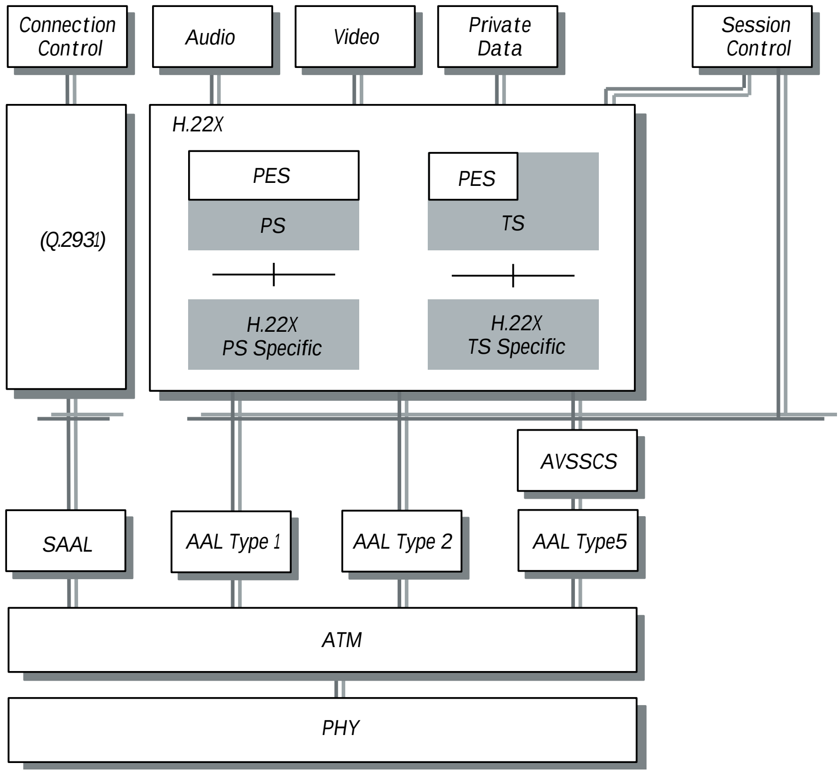


Figure 3.24: *ATM Forum baseline framework.*

without padding. Also, since no timing recovery is provided, the TS packets containing timing information may be subject to jitter, so the proposal is to packetize a single TS packet without any extra delay incurring in a loss of efficiency. This scheme is known as 5/8 or *PCR-aware*. The timing recovery was suggested to be done by an AVSSCS [139, 140].

Further investigations concluded that AAL5 was suitable for the targeted services. The reasons given were that the CDV should be bounded and below 1 ms which could be reasonably accommodated by the decoder given that the data stream is CBR. Actually, there is no proof that this CDV bound will always be met in real networks. Also the other QoS parameters should be guaranteed [141, 142]. In addition, the framework proposed in the baseline included the H.222.1 Network Adaptation layer and the AVSSCS that could perform the missing functions of AAL5 such as dejittering and synchronization. The development of an AVSSCS was discarded later because it was considered to provide the same functionalities as the MPEG-2 system layer. The synchronization was assumed to be done by the applications themselves and error correction functions were considered as a physical layer issue. Therefore, AAL5 was proposed for transport only [141, 143, 97]. This choice was completely unaligned with ITU-T and ETSI work which continued to support AAL1 and AAL2 for multimedia services [144].

The proposed solution was claimed to be a pragmatic short term solution due in particular to the large deployment of AAL5. By February 1995, release 1 of the Audiovisual Multimedia Services Implementation Agreement (AMSIA) baseline text focused only on VoD which was considered as the most urgent of the user requirements.

Once the choice was adopted in the first version of the implementation agreement, the discussion turned around the two possible packetization schemes, namely the 5/8 PCR-aware

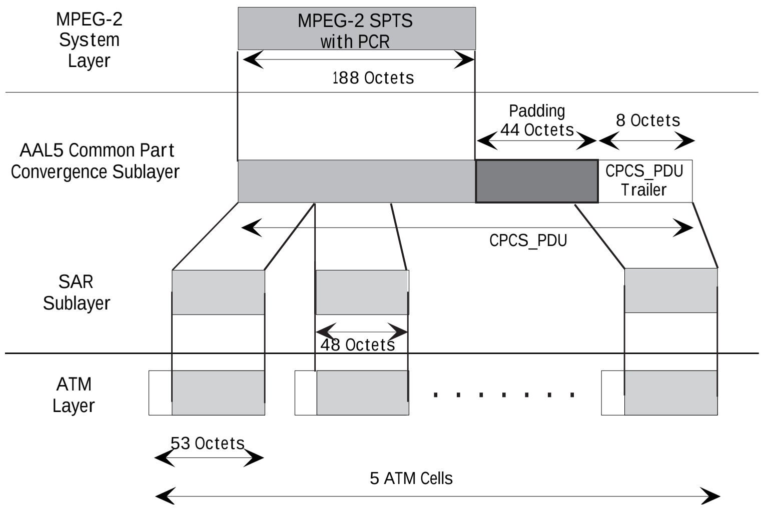


Figure 3.25: *PCR aware transport stream packing scheme.*

(Fig. 3.25) vs. the 8/8 PCR-unaware also known as *straight 8* (Fig. 3.26). Several issues had to be taken into account. First of all the target decoder. Whether a Set Top Terminal (STT) for the residential user or a general purpose desktop computer were the final user had important implications.

The 5/8 scheme has the advantage of reducing the jitter for the timing information but has the drawback of being more complex to implement given the fact that some layer has to be able to identify the PCRs in the TS which implies a knowledge of the data to be transmitted [145, 146, 147, 148].

Using a 8/8 scheme implied increasing the amount of memory needed at the decoder and therefore increasing the price [149, 150]. Bearing in mind that the driving force for the agreement was VoD, a mass market service, it was mandatory to develop a low cost STT. If a desktop computer had been the end-user target, then both 5/8 and 8/8 schemes had the disadvantage of generating small packets which implied moving a large number of packets several times per second, which today is not a task a computer has been designed for. Some contributions therefore proposed a default packetization scheme with the possibility of negotiating a larger PDU size (a 1/N scheme) [151]. Note that this discussion raised again the interest in using AAL1 because it has the advantage of not generating packetization jitter, like the 5/8 scheme, and also has the advantage of packing a single TS packet into 4 cells so it is neither necessary to packetize two TS packets per AAL5-PDU nor to be able to recognize the PCRs in the data, thus reducing simultaneously memory requirements and higher layers complexity [152, 153, 154].

The final implementation agreement was issued in January 1996 [155]. The document called Audiovisual Multimedia Services: Video on Demand Specification 1.1. addresses the carriage of MPEG-2 bit streams over ATM. In the first phase the specification addresses the requirements of VoD using Constant Packet Rate (CPR) MPEG-2 Single Program Transport Streams. The specification covers:

- AAL requirements
- the encapsulation of MPEG-2 Transport Streams into AAL5 PDUs
- the ATM signaling and ATM connection control requirements

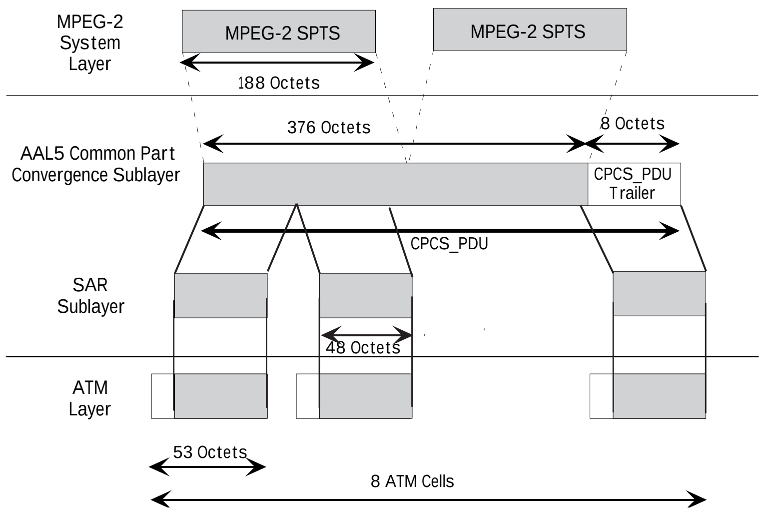


Figure 3.26: *PCR unaware transport stream packing scheme.*

- the traffic characteristics
- the QoS characteristics.

VoD is defined as an asymmetrical service to be used for entertainment purposes. The service will be predominantly point-to-point connecting a service provider, typically a video server, to a client, typically a STT or a PC. The user is able to select the video material and the time of viewing. Additional control functions such as *rewind*, *pause* and *fast forward* may also be available. However, the specification does not address user-to-user control which reduces the level of interaction to a minimum. Point-to-multipoint native services such as Near VoD (NVoD) or broadcast are not considered to be in the scope of the implementation since multipoint communications are not covered by the specification. The MPEG-2 video encapsulated in Single Program Transport Stream (SPTS) packets has to be mapped into AAL5 with a *null* SSCS. 1-to-N MPEG-2 TS packets are mapped into a single AAL5-SDU (PCR-unaware). However, all equipment must support a default mapping of $N = 2$. Concerning the handling of corrupted AAL5-PDUs it is said that if a packet is detected as corrupted with a correct length field then it is allowed to pass the PDU. Whereas this action is performed is left as an end station implementation option. This choice clearly means that bit errors are considered as harmless while cell losses are not. There is no evidence in the literature that this assumption is true.

Bearing in mind that the VoD specification targeted home services the economic factor was very important. The receivers containing the AAL and MPEG-2 decoders had to be low-cost. AAL5 is ubiquitous because it is required for signaling and it is simple to implement in hardware. Consequently, it is already produced in large volumes which is not the case for AAL1. The VoD specification which maes use of AAL5 benefits from this economy of scale situation. This implementation agreement clearly allows for a fast, large scale and economic deployment of a VoD service over ATM for the residential user. From the original set of services foreseen, the specification covers a simple service that has low real-time constraints.

By the time of writing, the ATM Forum SAA group has started AMS phase two which targets the transmission of VBR video. The major work items in the living list [156] are the following:

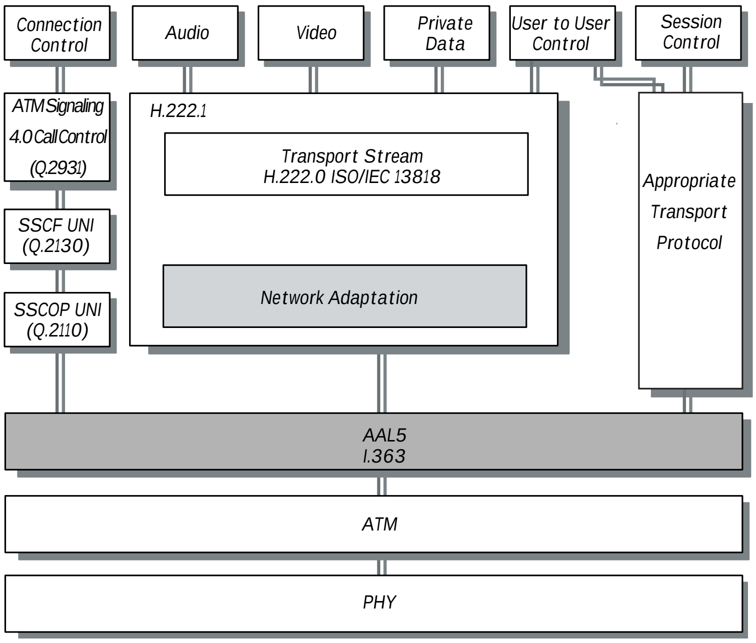


Figure 3.27: *VoD 1.1 user plane protocol reference model.*

- adoption of H.310 as the basis for the Broadband Multimedia Services work
- inclusion of Interactive Distance Learning and Multimedia Desktop services
- transmission of VBR encoded MPEG-2. Both, Program Streams and Transport Stream encapsulations should be considered
- choice of AAL5 for carrying VBR MPEG-2 data.

One of the driving forces beyond phase two is the development of the *Digital Video Disk*[1] (DVD). The characteristics of such device will be to use VBR encoded MPEG-2 encapsulated into PS packets. Although, originally DVD is designed to be used within a system in the same way CD-ROM is it seems that it could be used as a video source to be transmitted over broadband networks.

In addition to targeting this service, the options taken by the ATM Forum aim at evolving towards VBR services while guaranteeing backwards compatibility with AMS phase one. Moreover, the choice of H.310 aligns the AMS specifications to ITU-T recommendations.

3.5.3 Digital Audio Visual Council

The purpose of the Digital Audio Visual Council (DAVIC) is to advance the success of emerging digital audio visual applications and services initially of the broadcast and interactive type, by the timely availability of internationally agreed specifications of open interfaces and protocols that maximize interoperability across countries and applications or services. The goals of DAVIC are to identify, select, augment, develop and obtain the endorsement by formal standards bodies of specification of interfaces, protocols and architectures of digital audio visual applications and services.

[1] Recently, due to the many possibilities of such devices, the meaning of the acronym DVD has been changed to Digital Versatile Disk

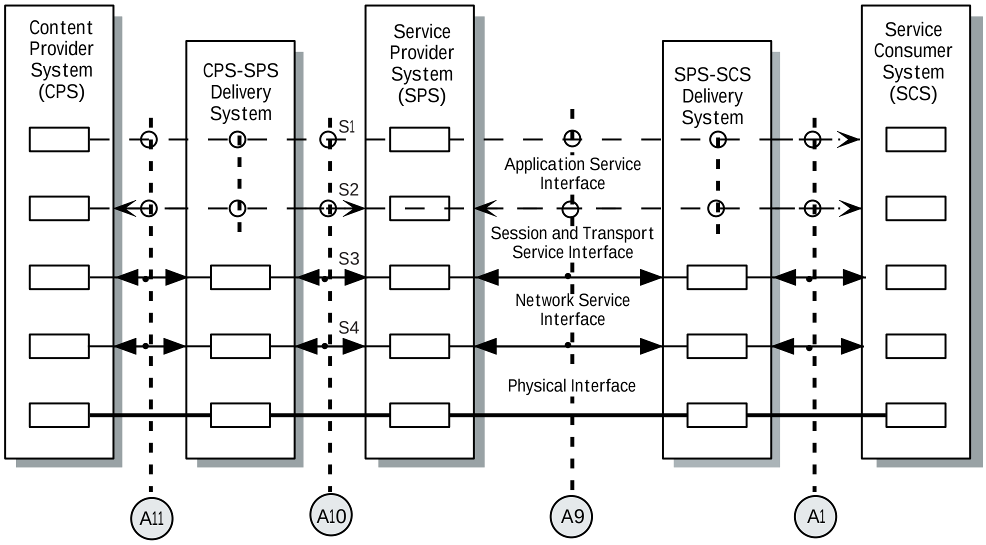


Figure 3.28: *The DAVIC framework.*

The framework of DAVIC specifications depicted in Fig. 3.28 is very general and aims at covering a wide range of multimedia applications.

It comprises three main entities: a Content Provider System (CPS), a Service Provider System (SPS) and a Service Consumer System (SCS). These three main entities are interconnected by two other entities: a CPS-SPS Delivery System which connects the CPS to the SPS and a SPS-SCS Delivery System which connects the SPS to the SCS. The DAVIC specifications can address any subsystem in this framework.

The DAVIC approach to specifications is to specify *tools* rather than systems which tend to be application specific. Conversely, tools are independent and may be easily relocated in any system.

DAVIC's intention is not to develop new standards for multimedia. Their approach is to make use of the existing standards to clearly define interfaces protocols and terminals to ensure interoperability between different equipments. Therefore, their proposals are aligned with ITU-T recommendations and ATM Forum specifications. Part 7 of DAVIC specification [157] which covers high and mid-layer protocols, defines a set of protocol components to support the different flows from a communication service (e.g. data, signaling, etc...) but rely on existing ITU-T and ATM Forum standards for transmission.

The standards cover the multipoint issue from the signalling point of view. They specify connection setup issues, but they do not specify the data transfer aspect of the problem.

3.5.4 Internet Engineering Task Force

The *Internet Engineering Task Force* (IETF) develops specifications for the internet. The development of the World Wide Web has given to the internet an unprecedented importance from an economical and informational point of view. This has also increased the interest for multimedia on the internet. Video conferencing software based on proprietary solutions flourish today. Also H.320 packages begin to appear. However, the increased functionalities available on the Web browsers are neither being followed by the network infrastructure itself nor by the protocols.

From a networking point of view, the increasing number of users leads to an almost permanent congestion state. In addition, the lack of guarantees for quality of service makes multimedia applications very difficult to deploy. An early solution, the Multicast backbone (Mbone), created an *overlay network* on top of the internet giving some dedicated resources, but still performance is extremely poor. From the protocol point of view, the TCP/IP protocols lack functionalities for real-time communications. Hence, new solutions are proposed to overcome these limitations. The Resource Reservation Protocol (RSVP) brings the concept of QoS and resource reservation to routed networks. Also the *Internet Protocol Version 6* better known as IPv6 has been designed to convey what is called a flow specification or *flowspec* which is a description of the characteristics of the data in order to reserve resources and therefore guarantee QoS. The concept of flowspec is also used in ST-II (see Sec. 3.3.4) which is an internet connection-oriented network protocol with resource reservation aimed at multimedia applications.

Concerning the transport protocols, the IETF has developed the Real-Time Transport Protocol (RTP) [4] which is an end-to-end network transport function suitable for applications transmitting real-time data such as audio and video, over multicast or unicast networks. RTP does neither address resource reservation nor any QoS for real-time services. It provides payload type identification, sequence numbering, time stamping and delivery monitoring. RTP relies on lower layers for multicast, error correction, timely data delivery and QoS in general. It has been designed to be independent of the underlying network layers. The main target of RTP is to satisfy the needs of multi-participant multimedia conferences.

RTP has been designed based on a new approach which follows the principles of application level framing (ALF) [158, 159] and integrated layer processing. ALFs principle is to explicitly include application semantics in the design of the applications protocols. The basic packetization has to be done at the application level. The advantage being that every packetization would be tailored to the specific application needs and would therefore be optimal.

RTP is intended to be configurable and flexible enough to be tailored to any particular application. Following this new approach, it will generally be integrated into the application rather than being implemented as a separate layer. Also, to achieve flexibility, the specification is deliberately not complete. It defines the core functions but it allows for extensions and customization depending on the target application.

RTP has originally been developed to work on top of UDP/IP but it may be used with other network and transport layers, in particular with ST-II. The basic functions that RTP provides are:

- multiplexing
- dejittering
- loss monitoring
- error detection
- payload identification.

The RTP header format depicted in Fig. 3.29 includes a sequence number to detect packet loss, a payload type identifier field, a time stamp that may be used for dejittering but also for delivery control via the RTP Control Protocol or RTCP.

RTCP, monitors the QoS and conveys information about the participants in a session. RTCP is not intended to support all of an application's control communication requirements. A higher level session control protocol, may be needed.

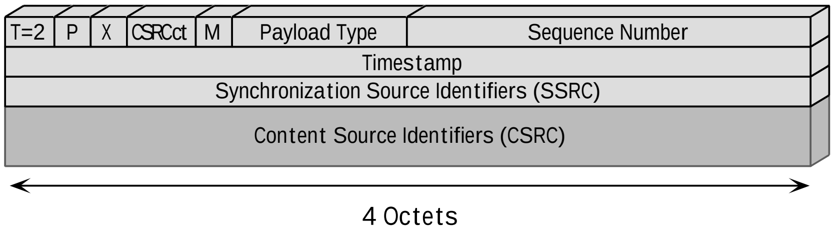


Figure 3.29: *RTP packet format.*

RTCP is based on a periodic transmission of control packets to all participants in a session, using the same distribution mechanism as the data packets. The underlying protocol *must provide* multiplexing of the data and control packets. RTCP provides four basic functions:

- Feedback on the quality of the data distribution. This is an integral part of the RTP's role as a transport protocol and is related to the flow and congestion control of other transport protocols.
- Persistent transport-level identifier for an RTP source called the *canonical name*. The identifiers are required by the receivers to associate the incoming flows to a given source and also to perform synchronization between multiple flows such as video and its associated audio.
- *Rate control*. In order for RTP to scale up to a large number of participants, RTCP provides means to calculate the number of users in a session. By having each participant send its control packets to all others, each user can independently observe the number of participants in a session. This value is used to calculate the rate at which each user may send data assuming that a session bandwidth has been allocated. To achieve fairness, all participants have to implement the same bandwidth allocation algorithm.
- Transmission of minimal session control information. This is an optional function that may be used for participant identification.

Since RTP specifies a set of core functions, companion documents have been written which specify the extensions needed for specific applications. In [160] the packet format for MPEG-1 and MPEG-2 audio and video is defined. They specify the payload identifier for the different packet formats (e.g. TS or PS). They also propose an encapsulation scheme which makes use of ideas already found in [75]. Some of the data stream headers are aligned to the beginning of the RTP packets.

If RTP functions are compared to what is proposed in H.222.1 the differences are minor. Both protocol layers propose multiplexing and an optional dejittering. None of them provide packet delineation. The network adaptation layer also proposes an optional sequence numbering and therefore a possible packet loss detection.

3.5.5 Summary

In summary, the standardization bodies have done little work to support VBR communications and in particular those with timing constraints. The current ITU-T and ATM Forum recommendations lack support for VBR interactive multimedia applications. AAL2, originally forseen for VBR, never reached an agreement and is now devoted to wireless ATM.

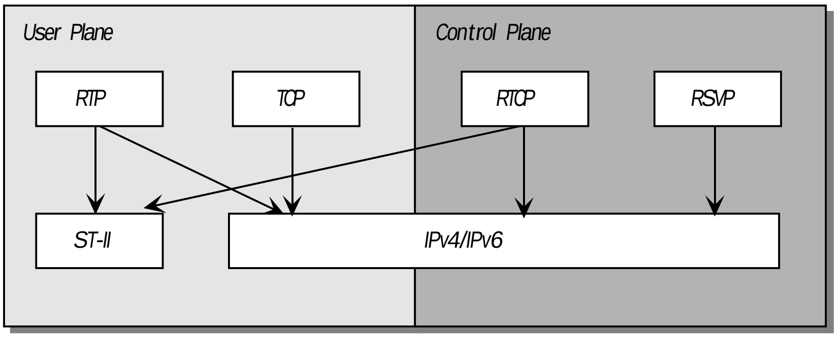


Figure 3.30: *Summary of discussed IETF protocols for multimedia.*

ITU-T's network adaptation layer currently does not support VBR traffic and the recent definition of the broadband multimedia terminal does include H.222.1 which restricts today's broadband terminals to support CBR services only and moreover it does not include AAL2 for future support.

This situation may change in some way. The ATM Forum already has work in progress to define the transmission of VBR video applications over ATM. However, the Audio Visual Multimedia Services phase 2 specification still relies on AAL5. Considering the alignment between ITU-T and the ATM Forum and the evolution of AAL2 towards wireless services, it is hard to believe that developments on a VBR-oriented AAL will be done in a near future. Moreover, ITU-T's endorsement of AAL5 for VBR interactive multimedia applications, definitely places AAL5 as the only choice for such multimedia applications, even if it was not originally designed for this purpose. This trend is clearly illustrated by the elimination of I.362 mapping of classes of service to ATCs.

3.6 Conclusion

Multimedia communications is a hot topic that still raises a lot of interest from the research community. Very much work has been devoted to the transmission of video over packet networks in general and broadband networks in particular.

To achieve a reliable transmission, from the application point of view, robust coding techniques which aim at reducing the impact of data loss by modifying the coding schemes or use special features to better fit to packet networks have been developed. Also, error concealment techniques to mask data loss at the decoder have been studied. Last but not least rate control which tries to control or adapt the application bit rate by different means.

The networking community has developed network and transport layer protocols for both broadband and internet networks to cover new issues like multipoint communications, resource reservation and QoS support in heterogeneous environments. To improve reliability, the study of FEC techniques for ATM has also captured a large of interest. Finally, the standard bodies such as the ATM Forum, ITU-T and DAVIC have been working towards the completion of a set of standards and specifications which will serve as the foundation to the fast deployment of multimedia applications.

However, almost all the efforts devoted to the transmission of interactive multimedia streams over ATM assume that one of the already existing ATM Adaptation Layers would be used. No layer has ever been defined for the transmission of real-time VBR data over ATM. Despite the fast development of multimedia communications and services, almost no work has been done to define the functions required by an AAL dedicated to VBR interactive multimedia.

Chapter 4

ATM Adaptation Layer for Interactive Multimedia Applications

4.1 Introduction

This chapter is devoted to the design and description of the services and functions proposed for a *new multimedia-oriented AAL*. We first present, based on the observations of the precedent chapters, a set of design principles we consider as mandatory for a *Multimedia AAL*, we will refer to as *MAAL*, capable of handling real-time interactive VBR multimedia applications in point-to-point as well as multicast environments. These design principles are then used as a foundation to derive a first set of functions for the MAAL which are:

- cell loss detection
- packet delineation
- dummy cell insertion
- cell loss correction.

Several options to implement each function are available. We discuss the different possibilities and present a rationale to every choice made. The first proposed implementation will be used as a basis for a performance study developed in Chap. 5.

4.2 Foundations for a New ATM Adaptation Layer

The foundations to build a new ATM Adaptation Layer are twofold; the requirements of real-time multimedia applications and the nature of compressed audiovisual data. The target of the precedent chapters has been the understanding of these two basic elements.

Chapter 2 is devoted to the description of the user requirements and the consequences they have onto the applications. We have also studied the syntactic and semantic structure of compressed audiovisual information. The available ATM network services are also reviewed and confronted to the applications requirements. We notice that among the different AALs none provides the network services tailored to what the requirements of general video dialtone applications are. AAL2 was foreseen as the layer that should have covered such services. However, it was never specified. Today, AAL2 has been devoted to wireless communications and therefore no further developments of an AAL for multimedia are to be expected. In addition, multipoint communications also require network services generally not available in ATM. Even if AAL3/4 provides a MID for multipoint connections, it has been banished from almost any utilization due to a too large overhead and complexity. Instead, the choice of AAL5 has profiled as the general solution because of its simplicity and most of all its ubiquity even if it does not provide the required features. Since its original target application was file transfer, it relies on upper layers for error correction mainly based on retransmission. However, the general belief is that multipoint communications require open-loop techniques for error detection and correction rather than closed-loop ones such as retransmission because, firstly it is difficult to handle such retransmission based error recovery techniques in high bandwidth-delay product networks, large buffers are required which in addition to round trip times generates an unmanageable delay. Secondly, such mechanisms tend to reduce the throughput due to flow control, not to forget the increased complexity that sources must have to handle large multipoint configurations.

It is still interesting to observe that when new services require AAL functions not yet covered by the standards, the ITU-T does not refuse new developments. We can therefore conclude that from a standards viewpoint real-time variable bit rate services do not need any further developments.

We described in Chap. 3 the vast amount of work that has been devoted to video dialtone services. Indeed, the transmission of real time multimedia data over ATM covers a large set of topics because the interactions between the user, the coding system and the network are extremely complex.

To achieve a reliable transmission, from the application point of view, robust coding, error concealment and rate control techniques have been largely studied.

The networking community has from its side extensively used queueing theory to characterize cell loss ratios under different types of input traffic. A lot of work has been devoted to study the performance of ATM multiplexers to derive Cell Loss Probabilities (CLP) and queueing delays. However, there has not been a main focus on the characterization of cell loss processes.

Recent contributions have shown that under the *low traffic source* condition which occurs when the fraction of the link rate used by a source is small, then the cell loss process could be approximated to a uniformly distributed process. The low traffic source assumption has been observed for periodic, On-Off and Gaussian sources.

The networking community has also developed new network and transport layer protocols for both broadband and internet based networks to cover new emerging issues like multipoint communications and resource reservation. To improve reliability, the study of FEC techniques for ATM has also captured a lot of interest.

Finally, the standard bodies such as the ATM Forum, ITU-T and DAVIC have been working towards the completion of a set of standards and specifications which will serve as the foundation for the fast deployment of multimedia applications. In spite of all the work done by the technical community, today only CBR based services could be deployed. The additional complexity that VBR data introduces has not been solved yet. To our sense, one of the reasons that drive this situation is the lack of an AAL providing such kind of services.

All these observations call for the development of an ATM Adaptation Layer that takes into account the requirements of multimedia applications. As a foundation for a *Multimedia AAL*, we decide to use the set of services originally defined by ITU-T in early drafts of recommendation I.363 [37] for the foreseen AAL2 which are the following:

- transfer of SDUs with a variable source bit rate
- indication of lost or errored information which is not recovered by the AAL

- transfer of timing information between source and destination.

This thesis, covers the first two services which are directly related to the end-to-end transfer of user data. The transfer of timing information which is related to synchronization and clock recovery issues are out of the scope of this work.

4.3 Structure of Compressed Audiovisual Information

Multimedia data sources are of diverse nature. Several compression algorithms exist with their own characteristics and syntax (see Sec. 2.1.2.1). Unlike multimedia, the data in file transfer applications does not have any particular syntax relevant to the network. All information is of equal importance and has to be transmitted without any error to the receiver. This is also the case for multimedia transmission today. All data is considered with the same level of importance even if solutions based on a combination of priorities and layered coding tend to prove that this is not the case.

Although each coder has its own syntax and semantics, some common elements exist that characterize all compressed audiovisual data streams, namely:

- the encapsulation scheme: audiovisual data is encapsulated into structures or packets delineated by headers. Since several levels of encapsulation exist, the information is hierarchically structured. This is due to the nature of the data which is also very structured (e.g. frames, lines, pixels). The second reason is the need that multimedia applications have to identify data streams of different nature which are multiplexed into a single bitstream

- the headers size: the size of the headers is always small. They can *always be encapsulated into a single ATM cell.* They represent a small fraction of the total data

- the sensitivity to loss: header losses have a larger impact than raw data losses. When a header is lost or damaged, none the underlying information can be recovered. The receiver's decoder loses synchronization and has to find a header of *the same or higher level in the hierarchy* to recover the synchronization. Conversely, when damage occurs in the semantic data, the decoder can, in some cases, still make use of the remaining data. The worst case being the recovery at the next resynchronization point

- the bit rate: compressed video sources generate data at a variable rate.

Since common characteristics between very different data syntaxes and semantics exist, we can envisage to define a set of *generic* functions that a MAAL should provide. The sensitivity and impact that data loss and headers in particular have, calls for a high degree of *reliability*.

This drives us to formulate a set of design principles we will use to specify the functions to be performed by the new AAL which are the following:

- genericity: the AAL has to support any type of codec

- reliability: errors or cell losses have to be minimized

- low delay: due to the nature of the data and the requirements of both the user and the application.

4.4 The Three Design Principles

4.4.1 Genericity

The AAL enhances the services provided by the ATM layer to support the requirements of a specific service, for instance the transmission of real-time variable bit rate applications. An AAL by definition has to be generic. By generic we mean that the AAL will not be codec-specific. It will only provide mechanisms for VBR multimedia applications in general. It has to remain generic in the sense that any type of compression algorithm may benefit from using such an AAL. However, to achieve the better possible service the nature of the data to be transmitted has to be taken into account. The difficulty is that several compression schemes will be used by multimedia applications and each codec has its own syntax and semantics. Still, some elements are common to compressed audiovisual data. As described in Chap. 2 the vast majority of compression algorithms organize the data in a highly structured and hierarchical way. This organization is based on a set of headers, the syntactic information, used by the decoders as resynchronization points that encapsulate the different data structures.

The way to take into account the nature of the data without developing an AAL per compression algorithm is to split the functions required into two sets: the network services to be provided by the AAL and application specific services to be provided by a *Network Adaptation Layer* (NAL) (see Fig. 4.1).

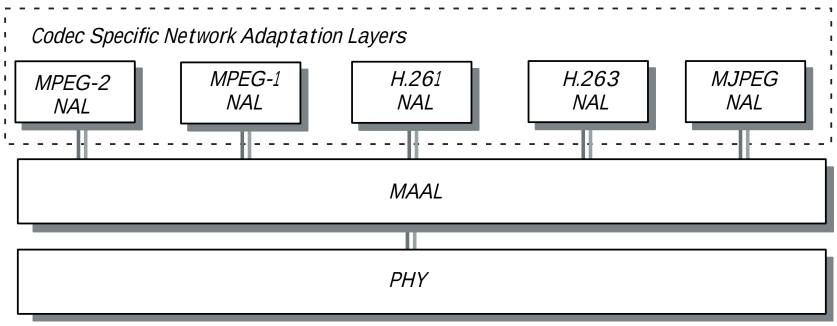


Figure 4.1: *Two layer reference scheme.*

Unlike a SSCS that could provide also service specific functions, the NAL will provide *codec* or *application* specific functions. The difference between service and application is that a service is defined as a set of primitives offered to the protocol user. Therefore, a service does not target a specific application but a group of applications with common characteristics. A SSCS could only provide functions for a type of service not for a specific application.

The reason why we choose this scheme is that, as it will be described in Chap. 6, to achieve efficient but low overhead protection schemes it is necessary to have a precise knowledge of the type of data to be transmitted in order to identify the sensitive information to protect. Bearing in mind that in multimedia applications *not all data is born equal* and also that some loss is tolerated, efficient protection schemes necessarily require such knowledge which requires the NAL to be codec specific.

Such a two-layered scheme provides application specific functions while keeping the network part, in this case the AAL, generic. This comes with the advantage that NALs could be used over other network or transport protocols if the underlying layers provide the required functionalities.

4.4.2 Reliability

In traditional communications a reliable transfer means a transfer of data without errors. In multimedia communications this is not a necessity since audiovisual communications are loss tolerant up to a certain degree. By reliable transmission, we do not envisage to guarantee an error-free transmission but a transmission where the errors have the smallest visible impact possible. In other words a reliable transmission is a transmission that *keeps the quality of service to a certain level independently of the network conditions*. This takes in account principally transmission errors.

Considering that interactive multimedia applications are of real-time nature, that they tolerate some loss and that, in addition, they will frequently be deployed on multipoint configurations such as the one of Fig. 4.2 we can easily derive the following observations: reliability via closed-loop techniques such as retransmission cannot be implemented because the delay will be at least equal to a round-trip time. This in many cases is not tolerable by the applications. Also, in large multicast configurations, the problem of feedback implosion arises if retransmission is envisaged. Therefore to achieve an error detection and correction mechanism tailored to both multipoint configurations and low latency requirements an *open-loop* mechanism is necessary. The mechanism which better fits to this scheme is the one based on *Forward Error Correction* (FEC) codes.

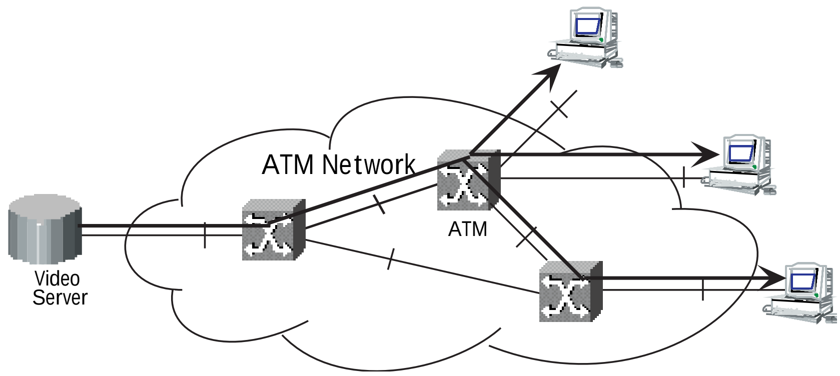


Figure 4.2: *Point-to-Multipoint configuration.*

One of the problems with FEC is that it does not guarantee the integrity of the data received. FEC codes cannot guarantee the recoverability of all the data. However, this is not necessarily a major problem in the case discussed here because multimedia streams and especially video tolerates some, yet limited, loss.

4.4.3 Low Delay

Obviously, an AAL providing functions for real-time applications has to be fast to satisfy the low latency requirement of such applications. The best way to achieve this is to implement simple mechanisms that do not trade complexity for delay. Concerning error correction open-loop mechanisms offer the smallest delay possible. However, if FEC has to be applied, it has to be done without the utilization of interleaving. Interleaving not only adds delay but moreover it is difficult to apply to VBR traffic. Also if we want to achieve short processing delays, FEC calculations must be kept to the minimum, therefore a simple FEC mechanism, such as an RSE based FEC (see 3.3.2.1), has to be used.

4.5 Proposed Functions for a Multimedia AAL

An ATM adaptation layer is composed of two sublayers. The *Segmentation and Reassembly* (SAR) sublayer segments the AAL-PDUs into 48 octets cells so called SAR-PDUs. These cells are then sent to the ATM layer which prepends the ATM header. It also performs the opposite function consisting of reassembling the incoming cells into AAL-PDUs. The *Convergence Sublayer* (CS) provides an AAL *Service Access Point* (AAL-SAP) to the layer above and is service dependent (see Fig. 4.3). The SAR and the CS sublayers may in some cases be empty. The convergence sublayer is further subdivided into a Common Part Convergence Sublayer and a Service Specific Convergence Sublayer (see Chap. 2 for a detailed description).

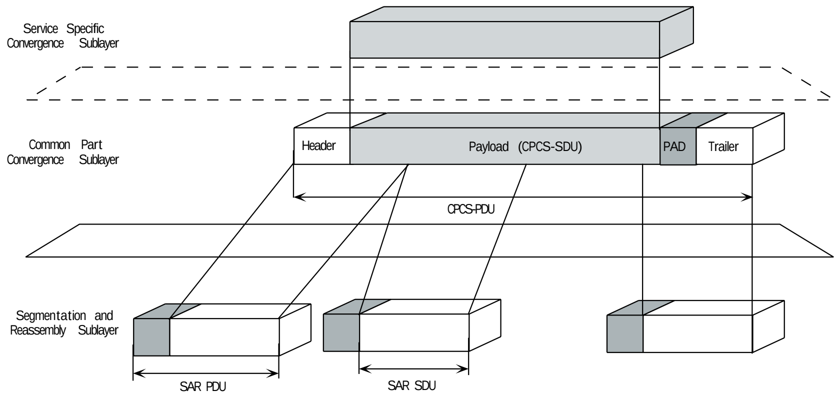


Figure 4.3: *General ATM Adaptation Layer structure.*

The functions that a multimedia AAL has to provide to support the aforementioned services are also taken from the early ITU-T specification of AAL2 [37]:

1. segmentation and reassembly of user information
2. handling of cell delay variation
3. handling of lost and misinserted cells
4. source clock frequency recovery at the receiver.

Among these four functions, this thesis proposes new mechanisms to *handle lost and misinserted cells* in depth, the *segmentation and reassembly* depending on the functions to be performed by the AAL. It also gives insights on how to perform the timing related functions.

We first address the handling of loss and misinserted cells function which covers:

- cell loss detection
- dummy cell insertion
- packet delineation
- cell loss correction.

The segmentation and reassembly depends on the SAR-PDU structure, which in turn is directly related to how the precedent functions are implemented, and will therefore be specified *a fortiori*.

The next sections propose a rationale for each of the AAL functions. To derive a first approximation required to evaluate the performance of the chosen functions, we make use of the low traffic source assumption [57, 58, 45]. The utilization of such a model is justified by the fact that high quality multimedia applications using compressed video should not make use of average bandwidths beyond 10 to 15 Mbits/s, even if peak rates could easily go beyond these values. We therefore assume that the cell loss process seen by this type of connection is independent and identically distributed which allows to develop some analytical models. These models are then used to justify our choices.

4.5.1 Cell Loss Detection

4.5.1.1 Cell vs. Packet Oriented Loss Detection

Cell losses will certainly occur in ATM networks due to different reasons such as congestion, transmission errors, network outages, UPC action, etc... Their frequency of occurrence will depend on several factors. Still, each AAL has to provide a mechanism to handle such a situation. How an AAL may react to cell losses depends on the class of service it was originally designed for. File transfer applications covered by class C and D services require a packet oriented AAL such as AAL3/4 or AAL5 because they are packet based applications. A reliable transport protocol such as TCP asks for the retransmission of a complete packet whenever it detects a corrupted or missing packet. It is therefore not necessary for the AAL to pass corrupted or incomplete packets since they will automatically be discarded by the higher layers and retransmitted. Besides, there is no need to add cell sequence numbering since the lost data position information is useless to the receiver and it adds an extra overhead.

Real-time applications such as voice and video, covered by classes A and B have a different behavior and nature. They are stream based applications. As such, they have delay constraints that in general make retransmission mechanisms inefficient. Moreover, audiovisual data tolerates some yet limited loss. In addition, audiovisual applications may take corrective actions if the errors are detected. It is therefore interesting to pass as much information as possible even if it is incomplete. This is the technique used by AAL1. It provides a cell loss monitoring capable of giving the number and position of the lost cells within the packet. In any case the corrupted packets are discarded since information valuable to the application may still be available in the received data. Instead, the packets are filled with dummy cells and passed to the higher layers with an error notification.

In summary, there are two ways of detecting missing cells: the first one consists in using a packet length information that is used to check if the length of the received packet is correct. The second one consists in using sequence numbers to monitor if there are no missing cells in the sequence. An advantage of the former packet-oriented mechanisms is that it generates low overhead. Note that AAL3/4 is seldom used, AAL5 being preferred instead, because it is considered to add too much overhead.

A packet-oriented layer using a packet length indication does not require sequence numbers which become redundant. Therefore, if we consider that passing corrupted packets (incomplete) is necessary they will be passed with an error indication but without any information concerning the position of the lost cells within the packet. It is therefore difficult to take advantage of the delivered information. Conversely, if a packet drop scheme is used, then the probability exists that the dropped packet contained a header that was correctly

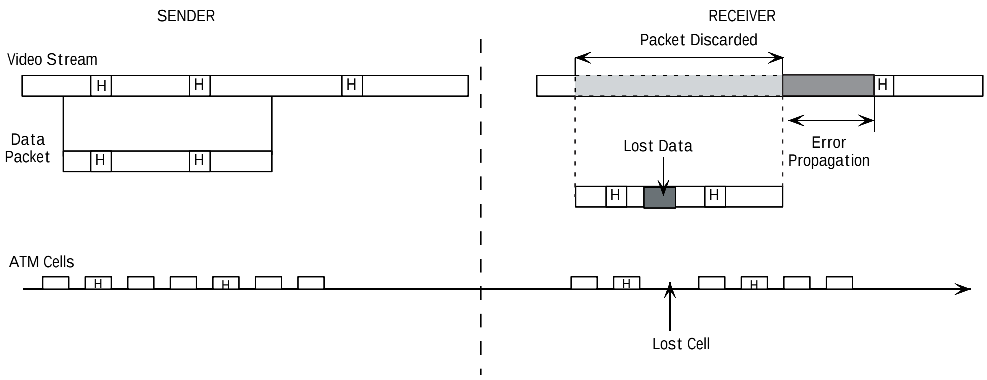


Figure 4.4: *Packetization of a video stream and loss propagation.*

received as depicted in Fig. 4.4. In this case, a resynchronization point would be lost amplifying the cell loss effect onto the video.

Instead, a cell based loss detection has several advantages. Among them the fact that a cell loss within a packet will not directly affect the remaining information of the packet because the corrupted packet could be passed to the upper layer with an error indication containing the number as well as the position of the lost cells in the PDU. This could easily be exploited by higher layers that could take corrective actions.

4.5.1.2 Sequence Number Field Size Dimensioning

The choice of a cell sequence number for the MAAL, requires that at least an octet from the SAR-SDU is used to carry the information. We propose here a preliminary study to determine the appropriate size of the sequence number *SN field* based on the low traffic source assumption.

A field size of n bits is able to number modulo 2^n consecutive cells. Once the maximum value is reached the counter is reset and numbering starts again from zero. This means that the AAL is able to detect up to $2^n - 1$ consecutive losses. It will however not detect losses of exactly 2^n cells or more. Therefore, the size of the SN field has to be carefully dimensioned in order to avoid, or reduce the probability of undetected losses to occur. To dimension the SN field one has to evaluate the probability of observing n consecutive cell losses.

Assuming a uniformly distributed cell loss process, we define the random variable X as the number of consecutive cell losses observed. Then if the cell loss ratio is equal to p the probability of observing n consecutive losses is given by:

$$P(X = n) = p^n.$$

Figure 4.5 shows the evolution of the probability for n between 1 and 8 bits. We have to consider that the assumption is the best case possible and that the real behavior would lead to higher probabilities of consecutive losses. We therefore chose $n = 5$, which allows to protect from bursts of up to 32 cell losses, assuming that no bit errors will occur in the sequence number field, as a preliminary value to be tested.

This leaves three extra free bits. As we will see in Sec. 4.5.4, the extra bits will be used for FEC information.

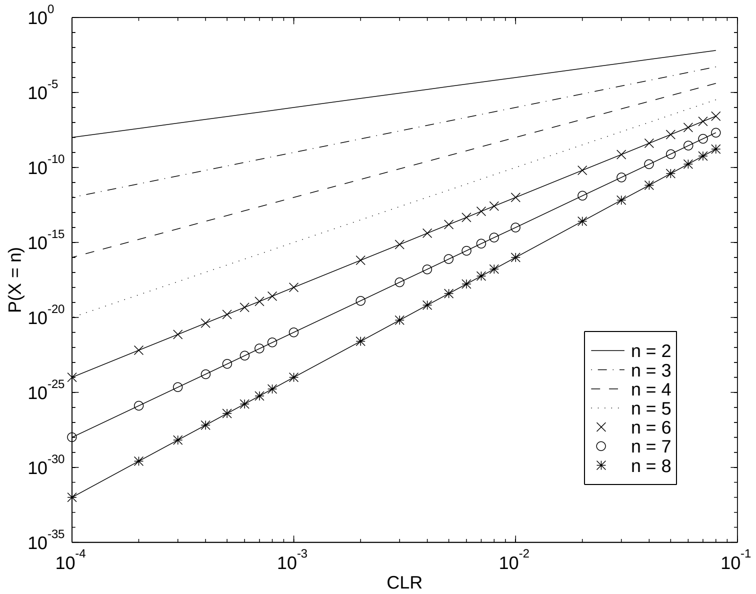


Figure 4.5: *Probability of losing* n *consecutive cells as a function of the CLR.*

4.5.2 The Dummy Cell Insertion Mechanism

A dummy cell insertion mechanism consists in inserting *dummy* data to replace the lost data within a packet. This mechanism could be used for different reasons. It could be used to keep packet size integrity. The reason why e.g. AAL1 inserts dummy cells is to keep the data aligned in the interleaver to allow recovery of the lost cells. Since AAL1 uses interleaving, if the data is not recovered, the lost data will be spread into several packets. The impact that this could have onto speech is almost not noticeable, however, for video, the situation could be much more different. There is no evidence in the literature that octet errors could be less harmful than cell losses. They could easily pass undetected through the decoder but still be very visible to the user.

If interleaving is not used then the only reason to use dummy cell insertion is to keep packet size integrity. It could be useful in the sense that if there is a packet length check at higher layers it could pass through, avoiding discard of the information.

We see a third possible utilization of dummy cell insertion related to video. Since all compression algorithms use run length coding, it is impossible that long sequences of the same value appear. Therefore the existence of an unallowed codeword could be interpreted by the decoder as an error which could then drive the error concealment mechanisms.

Taking into account that different types of applications and codecs could make use of the MAAL, we propose the dummy cell to be *selected by the user* at connection setup.

4.5.3 Packet Delineation

If variable packets are used, it is necessary to provide a packet delineation mechanism. The ATM cells have a 2 bit field in the header, the Packet Type Identifier (PTI), whose purpose is to delineate the packets. The PTI field of the first cell of a PDU is set to Beginning of Message (BOM), the intermediate packets to Continuation of Message COM and the last

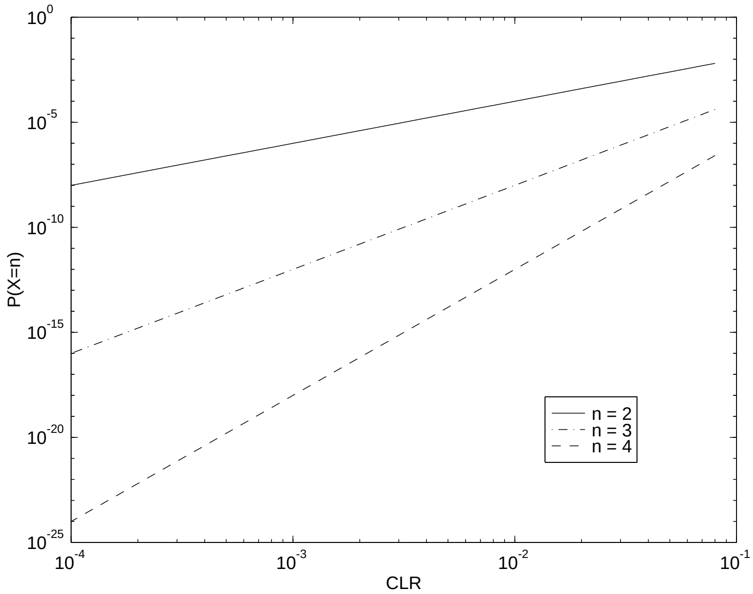


Figure 4.6: *Probability of consecutive packet losses due to the loss of delineation cells as a function of the CLR.*

cell of the PDU is set to End of Message (EOM). If only these fields are used to delineate the packets then the probability exists that the delineation cells are lost leading to errored packets reassembled and transmitted to the higher layers. For this to occur, both the EOM and the next BOM cells have to be lost. A first evaluation could be done by calculating the probability that this happens assuming that the cell loss process is uniformly distributed. Then, the probability of losing a cell is independent and identically distributed. Given a cell loss ratio of p, the probability of losing a BOM or an EOM is the same and equal to the probability of losing any other cell. The value does not depend on the size of the packet. Also, the probability of losing both an EOM and the consecutive BOM is also independent. Thus, if we define $P(A)$ and $P(B)$ as being the probabilities of losing an EOM and a BOM respectively, the probability of losing the delineation cells for packets of size m is equal to the probability of losing any two consecutive cells:

$$\begin{aligned} P(A \cap B) &= P(A) \times P(B) \\ P(A \cap B) &= (p)^2, \end{aligned}$$

and does not depend on m. The probability X of observing n consecutive packets, and therefore $n \times m$ cells, lost due to loss of delineation cells is given by:

$$Prob(X = n) = \left(p^2\right)^{(n-1)}.$$

As Fig 4.6 shows, the probability of multiple consecutive packet losses due to loss of delineation packets is small. This means that under the low traffic sources assumption, *the utilization of the PTI fields in the ATM cells should be enough to reliably delineate and transmit variable size packets.*

The disadvantage of transmitting variable size packets is that they do not necessary segment into an integer number of SAR-SDU payloads. Therefore, the CPCS has to provide a boundary alignment function. To align CPCS-SDUs to SAR-SDUs, byte stuffing or padding is required. If padding is performed then extra information has to be carried to the receiver's end to indicate the number of extra bytes transmitted. This requires a CPCS-SDU header or trailer that will necessarily add overhead.

4.5.4 Cell Loss Correction

One of our design principles for the AAL is the reliability. The AAL must handle information which tolerates some loss, that will make wide use of multipoint configurations and most of all has a structure with small headers of very important information. These reasons make us believe that the utilization of an open-loop error correction technique is the best solution. Forward Error Correction seems to fulfill all the requirements of multimedia applications and as already discussed in Sec. 3.3.2.2 using it in ATM has been the subject of several publications. However, as already mentioned, one of the drawbacks of FEC is that it does not necessarily recover all the data lost. Biersack shows in [84] that when used in ATM networks FEC efficiency depends on several parameters and in particular in the cell loss pattern. Correlated or bursty loss patterns make FEC inefficient. Also, the density of bursts has a strong influence on the recovery ratio of FEC.

Another largely debated question is the position of the error recovery functions within the protocol stack. Two tendencies exist: one consists on correcting errors as close as possible to the application where the data has to be delivered to the user while the second tries to correct the errors as close as where the errors occur as possible. The first case is the one we find today. The strategy in fact is to leave all error correction functions to the decoder itself. The advantage of this technique is that no extra functionalities are required. The second case is the one we propose. Several reasons, in fact, argue in favor of the cell level error recovery; the loss tolerance of the applications, the open-loop technique which leads to low delay, the granularity of the data and the fact that error recovery is more efficient when it is performed close to where it occurs.

We describe in the next sections the different recovery techniques based on FEC and their efficiency. We show the difficulties associated to the granularity of frame based FEC and the long delays it may generate. We show that the AAL level FEC is flexible enough and encompasses real-time multimedia requirements including VBR applications. We also show that it is possible to overcome the efficiency reduction due to loss correlation with non-interleaved techniques.

4.5.4.1 Forward Error Correction Mechanisms

The trend for ATM based multimedia applications is to leave error correction functions to the application. The main reason for this choice is because AAL5 is used to transmit multimedia streams [161] and AAL5 does not provide enough functionalities to perform error correction. Several papers however [83, 162], advocate for the utilization of FEC in real time multimedia communications.

Basically, there are three main techniques to apply FEC to packet networks. All these techniques are based on Reed-Solomon error correcting codes (RSC) or on a simplified version able to correct erasures only called Burst Erasure correcting code (RSE) [83]. Both RSC and RSE codes, introduced in Sec 3.3.2.1, may use or not interleaving byte or cellwise.

We discuss in this section the characteristics of the three techniques.

4.5.4.1.1 Reed-Solomon Octet Interleaver

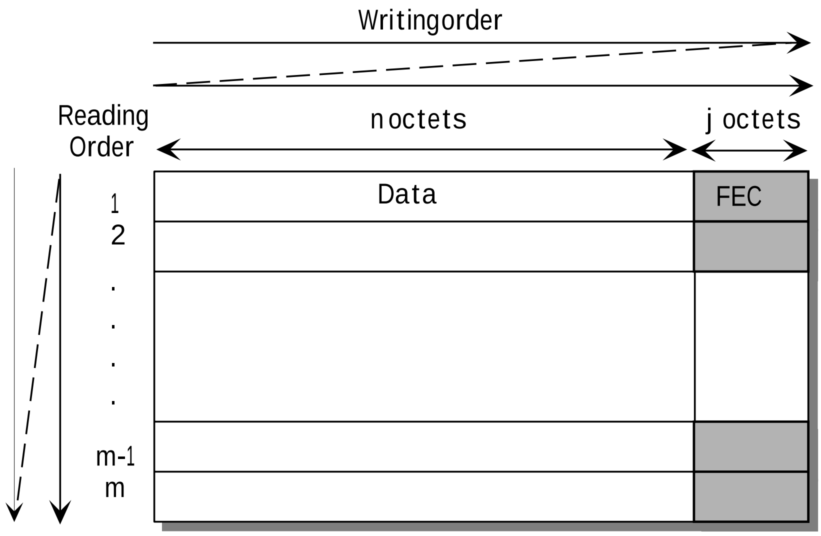


Figure 4.7: *Reed-Solomon based long interleaver.*

The implementation of FEC on ATM, introduced in Sec. 2.2.3.1, has already been studied and implemented in AAL1 [37]. It provides a CBR based service with timing recovery. The main applications to use AAL1 are circuit emulation applications or in other words telephony over ATM. It is well known that speech is very sensitive to losses due to its temporal coherence and therefore needs error recovery methods. Moreover, the transmission of voice is a real-time service with very stringent timing constraints. Therefore FEC has been adopted to achieve low loss ratios and low delays. The FEC implementation used in AAL1 is like the one depicted in Fig. 4.7. This mechanism uses a combination of Reed-Solomon codes with interleaving. The parity data is used to recover the errors while the interleaving decorrelates loss bursts. This combination allows this code to recover from erasures (losses) and also from random errors (impulse noise). If we consider that the size of a column read bytewise is equal to a cell payload of m octets, the performances obtained for such scheme applied to the ATM context are:

- cell loss recovery: $j = h\ cells$
- cell/octet error recovery: combinations of i cells lost + l octet errors with $i + l \leq j$
- octet errors: $\frac{j \times m}{2}$ errored octets per row if there is no cell loss
- FEC overhead: $\frac{j}{n+j}$
- end-to-end delay: $D_{sender} + D_{receiver} = 2 \times (n + j)\ cells.$

This code is very powerful in terms of error recovery. However, it is worth to note that the ratio of bit errors expected in fiber optics based networks is in the order of 10^{-12}. In fact, with such impulse noise error values, it becomes unnecessary to have bit or octet granularity. Therefore, and taking into account that it introduces considerable delay, an octet based interleaver is not a major need. A different interleaver method also specified by ITU-T allows to reduce the end-to-end delay. It is based on a short interleaver where the data is

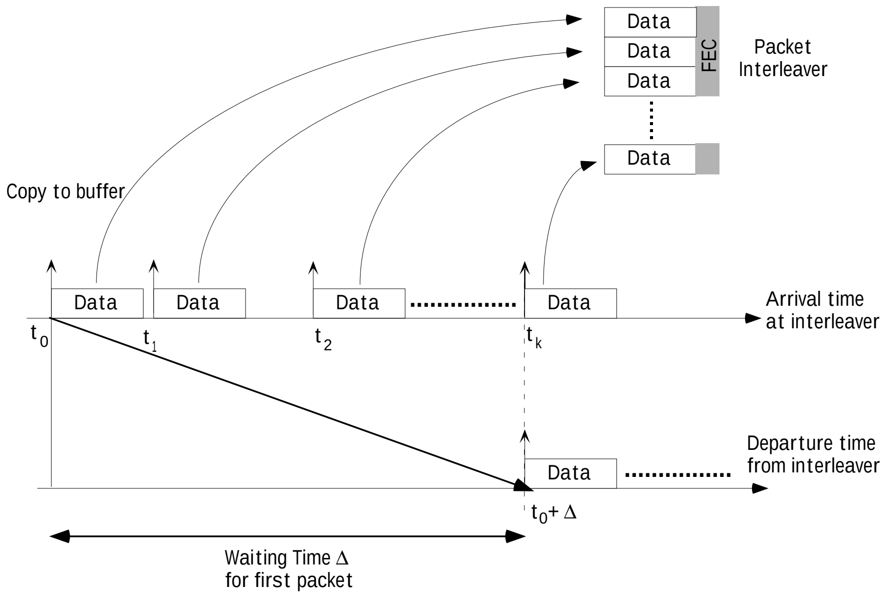


Figure 4.8: *Packet interleaver delay.*

read diagonally rather than top to bottom from the matrix. However, this method increases the overhead by a factor of 2 while reducing the correction ratio by a factor of 4.

If this technique is to be applied to VBR communications several drawbacks appear. First, the mechanism relies on a fixed data matrix structure defined by the interleaver. If we refer to Fig. 4.8 it is easy to see that each of the packets P_i will have a different waiting time, the longer being for the first packet and the shorter for the last one as follows:

$$\begin{aligned} \Delta\ t_{P_1} &= t_k - t_0 \\ \Delta\ t_{P_2} &= t_k - t_2 \\ &= \vdots \\ \Delta\ t_{P_k} &= 0, \end{aligned}$$

assuming a null processing delay.

Since the arrival times for VBR are unknown the waiting times for each of the packets, but the last, are also unknown.

Even if this is manageable for CBR applications since the jitter introduced is constant, it may become a problem if this technique is applied as is to VBR applications, since the jitter experienced by each of the packets will depend on its arrival time to the interleaver, time that is not known by the receiver. Another problem of this scheme is that given the fact that it is an octet based mechanism it is necessary to have a fixed structure in particular for the number of rows. To accommodate packets, either they have to be segmented into several rows which will probably need padding or they have to be of fixed size. The AAL1 implementation uses 47 rows of 124 octet packets to build up the ATM cell payloads.

Nevertheless, an advantage of this mechanism is that with small effort, the overhead could easily be modified by increasing or reducing the number of parity bytes per line. It is therefore possible to envisage selective or variable protection schemes if, howsoever, the

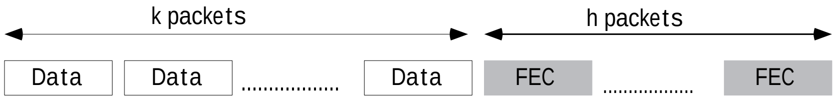


Figure 4.9: *RSE based FEC.*

jitter problem could be solved.

4.5.4.1.2 Non-Interleaved Burst Erasure Protection

The second method (Fig. 4.9) is based on burst erasure codes (RSE). This method takes into account the specifics of ATM which are twofold: erasures are limited to fixed boundaries and bit errors are negligible. Both conditions are true in ATM networks. Burst Erasure codes use the same polynomial algebra to generate the overhead than RSC. However, unlike RSC, they are unable to correct random losses due to impulse noise. Without interleaving, the correcting capabilities of RSE are reduced to erasures only. However, delay can be considerably reduced for the same reason. With this scheme, the data packets are sent in the same sequence as they are received from the upper layer. So, if a copy of the packet is held at the sender, the data can be sent prior to the calculation of the FEC data. This reduces the end-to-end delay by a factor of 2 (see Fig. 4.10). A second advantage is that overhead can easily be modified by adding supplementary FEC cells (e.g. overhead proportional to the frame size). Also, not having a fixed matrix structure, makes this scheme suitable for VBR traffic since it will not add any jitter due to accumulation of data in the interleaver since the packets are going to wait the same *constant* processing time before being sent. Finally, in terms of correction, this mechanism is as efficient as an octet based interleaver. It can correct any combination of losses as far as the receiver gets any k out of the $h + k$ packets sent, h being the number of redundancy packets. RSE efficiency is directly proportional to the number of redundancy packets applied.

Last but not least this mechanism could easily be applied at the cell as well as at the packet level without modification. For sake of simplicity, we compare the different mechanisms applied at the cell level.

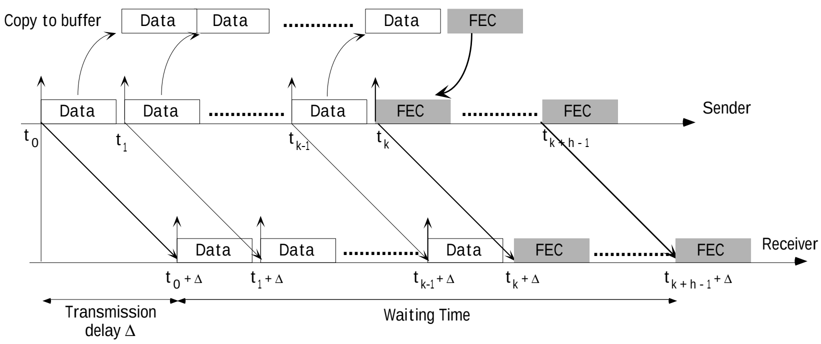


Figure 4.10: *RSE End-to-End delay.*

To summarize the performance of this method:

- cell loss recovery: h *cells* (any combination of h out of $h+k$)
- no random loss (octet errors) recovery, detection possible

- FEC overhead: $\dfrac{h}{h+k}$
- end-to-end delay: $D_{receiver} = h + k\ cells$.

4.5.4.1.3 Burst Erasure Cell Interleaver

The third method already proposed in [163] uses a combination of both techniques. The principle is to take advantage of the interleaving technique to cope with burst losses while keeping a cell level granularity. It makes use of RSE to generate the redundancy data cells and of a cell based interleaver.

Performances of this mechanism have been studied in [163] where a comparison between non-interleaved and this interleaved method are done. The conclusions derived are that for VBR in particular, the interleaved method is by far more efficient than a non-interleaved method concerning the CLR albeit this method is not as efficient as the other two to correct erasures. If we refer to Fig. 4.11 it is easy to see that if more than j cells are lost in a single row the redundancy cells are not able to recover the data in the row. This means that the receiver is not able to recover all the data under the condition that it gets k cells out of $h + k$ cells sent where $k = m \times n$ and $h = m \times j$. However, the error recovery capability is closely related to the delay and it is not clear whether it is better to have a small loss ratio with some delay or conversely. It is worth to say that one of the key parameters that may determine efficiency of one or the other methods is the frame sizes. For large frames an interleaved method is far more efficient while if small packets have to be protected low delay techniques may be enough.

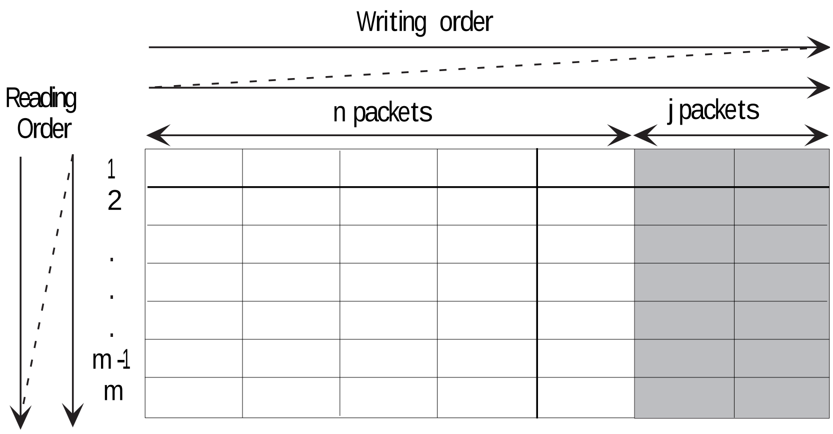


Figure 4.11: *Cell based interleaver.*

This technique gives the following performances:

- cell loss recovery: $m \times j\ cells$ if less than j cells in a single row
- no random loss (octet errors) recovery, detection possible
- FEC overhead: $\dfrac{j}{n+j}$
- end-to-end delay: $D_{sender} + D_{receiver} = 2 \times\ (n + j) \times m\ cells$.

Table 5.4.3 summarizes the delays and error recovery performances obtained for each of the three methods. To make the comparison as fair as possible we assume:

- the size of a packet is equal to m cells
- overhead is constant
- the matrix size of the cell based interleaver is given by: $\frac{k}{h}$ *packets per row*, which means that we try to have a single FEC packet per row.

	cell recovery	octet recovery	overhead	Delay (cells)
RSC Octet Interleaver	h cells	$i+l \leq j$	$\frac{h}{h+k}$	$2 \times (h+k)$
RSE non-interleaved	h cells	none	$\frac{h}{h+k}$	$h+k$
RSE Cell interleaver	$m \times h$ *cells*	none	$\frac{h}{h+k}$	$2 \times \ (h+k) \times m$

Table 4.1: *Summary of FEC based error correction performances.*

From this table we can conclude that the RSE based FEC mechanism without interleaving gives the best delay/error recovery trade-off. Moreover, it has the advantage of being flexible enough to handle VBR data.

In the next section we discuss how these mechanism may be implemented and how it would perform at the cell or packet level.

4.5.4.2 Packet Level Forward Error Correction

AAL5 was developed mainly for data transfer applications but nowadays it is considered as a general purpose AAL. It is simple to implement and generates low overhead. Today all the Network Interface Cards (NICs) that can be found in the market have AAL5 built-in. On the other hand, not all the NICs have AAL1 or AAL3/4 available. It would be therefore an advantage to have a frame based error protection mechanism that could be built on top of AAL5 able to work on every NIC.

Working on top of AAL5 gives a packet visibility to the error correction mechanism. This presents several drawbacks. The first is related to delays. If we consider the delay formulas in table 5.4.3 we have to consider that the values are multiplied by the number of cells per packet. The second problem is related to the error recovery efficiency. If we consider the mechanism of Fig. 4.9 we are protecting k packets with h overhead packets. To guarantee no loss it is necessary to receive k out of $k+h$ packets. For this condition to happen, it is necessary to lose less than h packets. Considering that losing a cell in a packet implies the loss of the packet, if $h+1$ cells are lost each one in one packet then we cannot recover the $h+1$ lost packets. This is a worst case situation for AAL5 that occurs if the losses observed in a single connection tend to be non-correlated, in fact, close to a uniform distribution.

To prove the error recovery efficiency reduction for the packet level scheme lets assume an independent and identically distributed cell loss pattern. The loss process is, in this case, of type Bernoulli. If we define the random variable A as a cell loss event then:

$$\begin{cases} P(A) & = & CLR \\ P(\bar{A}) & = & 1-CLR \end{cases}$$

Consider k packets protected with h FEC packets as shown in Fig. 4.9. Each packet segments into m cells. We want to calculate the probability of receiving at least k among $k+h$ correct packets.

The probability p of losing a packet is given by the probability of losing at least one and at most all the cells in a packet if we consider that AAL5 is used. We want to calculate the number of successful events (cell losses) of a Bernoulli process. Thus the probability of losing a packet is given by a binomial distribution.

$$p = \sum_{n=1}^{m} \binom{m}{n} CLR^n \times (1 - CLR)^{m-n}. \tag{4.1}$$

As Fig. 4.12 illustrates AAL5 amplifies the CLR by a factor equal to m. Therefore a solution to reduce the number of packets lost due to cell losses is to use *small packets*. This reduces the *packet discard* effect of AAL5. Note that beyond the fact that AAL5 increases the CLR seen by the application, data that has been correctly transmitted and could still be used by concealment mechanisms, such as speech interpolation or early resynchronization, is dropped by the receiver. It is worth to note also that a packet based recovery mechanism will not benefit from I.363.5 recently added feature [127] that allows AAL5 to pass corrupted PDUs. Indeed, this feature allows to pass corrupted packets that have a correct length (i.e. no cell losses). Moreover, such a mechanism is incompatible with error recovery techniques because using corrupted packets to recover missing packets will lead to wrong results.

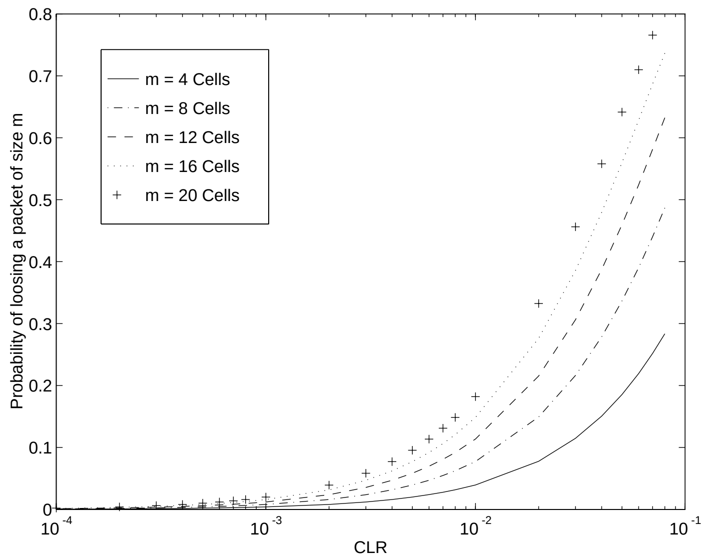


Figure 4.12: *Probability of losing a packet of size m cells as a function of the CLR.*

If the cell loss process is *iid*, then the packet loss process is also a Bernoulli process with a packet loss probability p given by equation 4.1. If we define the random variable X as being the number of received packets, the probability $P(X \geq k)$ of receiving k among $k+h$ packets is again given by a binomial distribution:

$$P(X \geq k) = \sum_{n=k}^{k+h} \binom{(k+h)}{n} (1-p)^n \times p^{k+h-n}. \tag{4.2}$$

Figure 4.13 shows $P(X \geq k)$ based on Eq. 4.2. We have fixed $k = 8$ packets and $h = 1$ packet. The packet size has been varied between 4 and 20 cells. This protection scheme allows a single packet to recover one lost packet which means that a single packet loss only is tolerated regardless of the number of cells lost *within the packet*. This scheme is therefore sensitive to the distribution of the cell losses. With a bursty loss process, which increases the probability of consecutive losses, such a mechanism would perform better than under the uniformly distributed assumption.

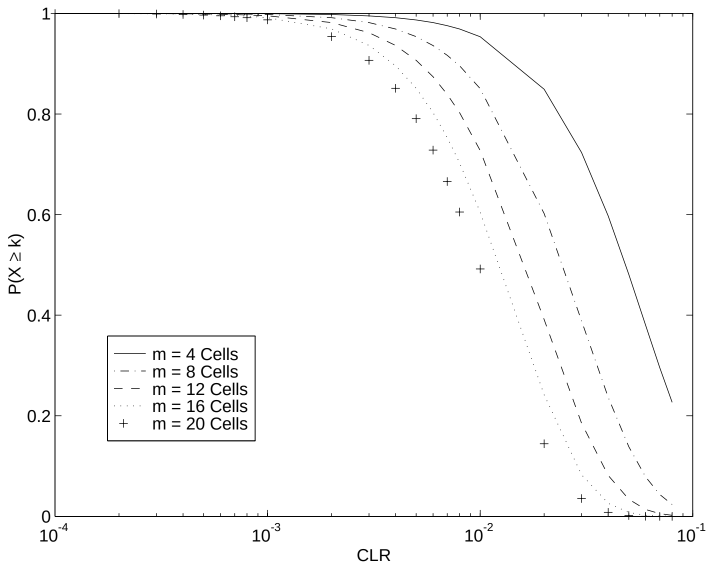


Figure 4.13: *$P(X \geq k)$ for packet level protection as a function of the CLR.*

Figure 4.13 depicts $P(X \geq k)$ for different packet sizes. This mechanism is very sensitive to the packet size. For a given CLR the probability of losing a packet increases according to Eq 4.1. The recovery efficiency is reduced by a factor m as shown in Fig. 4.13.

4.5.4.3 Cell Level Forward Error Correction

If we now consider that the FEC data is generated at the cell level, in this case the SAR sublayer, to achieve the same overhead as in the packet level protection case we have to send $k \times m$ cells that are protected by $h \times m$ FEC cells. Then, using the same random variable X as in Eq. 4.2, the probability $P(X \geq k \times m)$ of receiving $k \times m$ cells among $(k+h) \times m$ is given by:

$$Prob\{X \geq (k \times m)\} = \sum_{n=mk}^{m(k+h)} \binom{m(k+h)}{n} (1 - CLR)^n \times CLR^{m(k+h)-n} \tag{4.3}$$

Figure 4.14 shows $P(X \geq k \times m)$ for the cell level recovery mechanism according to Eq. 4.3. We keep the same k, h and m parameters of the packet case. In the cell level scheme, all the redundancy cells protect all the data cells and therefore recovery depends only on the number of lost cells and not on the distribution within the *FEC block* which explains the improved efficiency compared to Fig. 4.13. Note that increasing the packet size *increases* the recovery efficiency of the cell level mechanism. This is due to the fact that for a given CLR, the probability of losing m cells is a function of the CLR while the probability of receiving $m \times k$ cells is a function of Eq. 4.3.

4.5.4.4 Comparative Performance

Undoubtely, the cell level protection achieves a much better performance. The recovery ratio is 100% for CLR values up to 10^{-2} for all packet sizes. The packet level FEC begins

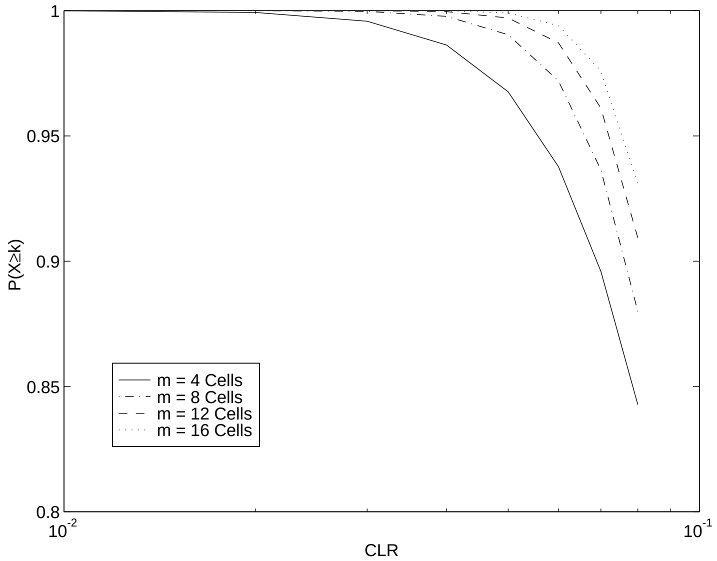


Figure 4.14: $P(X \geq k \times m)$ *for cell level protection as a function of the CLR.*

to lose data for CLR values around 10^{-3}. The cell level protection scheme performs more than an order of magnitude better for the same *protection overhead*. Note that the packet based protection heavily depends on the packet size. The larger the packets, the higher the probability of having more than one corrupted packet which in turn reduces recovery efficiency. In summary, the efficiency of the packet level protection scheme depends on:

- cell loss process
- packet size

and the efficiency of the cell level protection scheme depends on:

- number of redundancy cells.

This gives a clear advantage to a cell level protection scheme.

Figure 4.15 shows a comparative plot of the recovery efficiency achieved for both packet and cell level schemes. In Fig. 4.15 (a) we have fixed for the packet level scheme:

- $k = 8$ packets
- $h = 1$ packet,

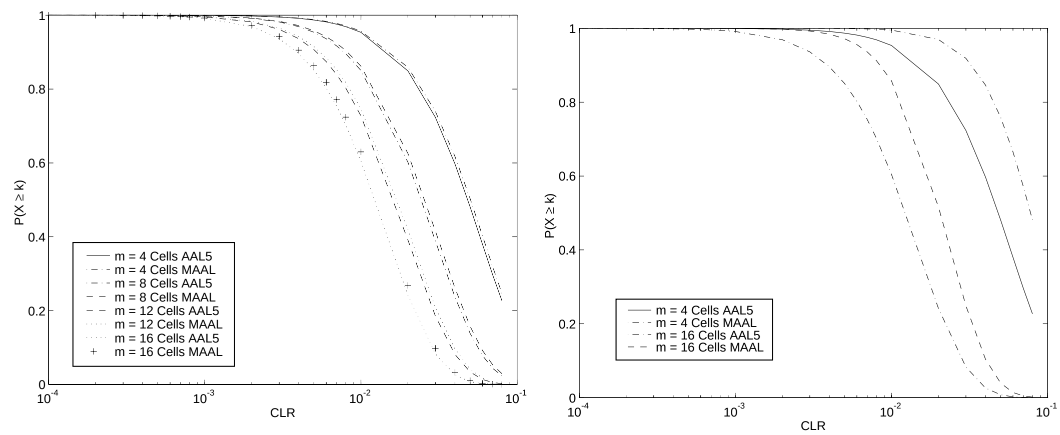
and for the cell level scheme:

- $k = m \times 8$ cells
- $h = 1$ cell,

where m is the number of cells per packet.

The cell level mechanism achieves the *same recovery efficiency* as the packet level with an overhead m *times smaller* than for the packet based case for all the packet sizes considered.

Figure 4.15 (b) shows again a comparative plot for both protection mechanisms. In this case the cell level overhead is $\frac{2}{m}$.



(a) Packet-based FEC

(b) Cell-based FEC

Figure 4.15: *Comparative packet and cell based protection probability of receiving k among $k+h$ packets as a function of the CLR.*

Another advantage of the cell level protection is its level of granularity. The overhead is much easier to manage at this level than at the packet level. If we define the overhead as being:

$$\begin{aligned} overhead_{packet} &= \frac{h}{h+k} \\ overhead_{cell} &= \frac{n \times h}{n \times h + m \times k}, \end{aligned}$$

at the cell level it possible to use the equivalent to a fraction of a packet size which is not possible at the packet level. For example, at the cell level it is possible to protect k packets, or $m \times k$ cells with r cells, with r not necessarily being equal to $m \times h$.

In summary, and under the uniformly distributed cell loss process assumption derived from the low traffic source hypothesis, a cell level protection mechanism achieves a much better performance than a packet level one. We therefore propose the MAAL to provide a cell level RSE based FEC. To improve recovery performance, we also propose to limit the PDU length to small sizes. This mechanism has the advantage of providing a recovery performance which depends only on the number of redundancy cells transmitted. It is much more flexible in terms of overhead, due to its finer granularity and provides the best delay/recovery trade-off of all the presented data recovery mechanisms.

4.5.5 Proposed Mechanisms

We derive in this section a first set of functions for the MAAL based on the previous analysis. We propose the MAAL to be a cell-oriented AAL. The functions the MAAL will perform are the following:

- cell sequence numbering
- transmission of constant size packets

- passing of corrupted PDUs to the upper layers with an error notification
- user selected dummy cell insertion mechanism depending on the type of application or codec
- selective RSE based Forward Error Correction.

To achieve a reliable transmission it is necessary to have an efficient cell loss detection mechanism. This could only be achieved if the AAL provides a cell sequence number. This requires from the SAR sublayer to add extra information in a cell header or trailer. Since any data is at least byte aligned, the minimum size for a SAR-SDU header or trailer is one octet. This reduces the payload size to 47 octets therefore adding overhead. However, we have seen that sequence numbering has several advantages. It allows to perform dummy cell insertion. Also, the size and position of lost data could be very useful to the application if error concealment mechanisms are to be used. Finally, if an RSE based cell level recovery mechanism is to be implemented, a cell sequence numbering becomes mandatory.

The transmission of variable size packets is possible because using the available PTI fields only does not penalize the transmission in terms of overhead. We have shown, under certain assumptions, that the probability of packet loss due to the loss of delineation cells is almost negligible. However, variable size packets require extra functions from the CPCS. Whenever data packets are not an integer number of SAR-SDU payloads, padding is required which entails extra overhead. If we take into account that some compression algorithms such as MPEG-2 deliver constant size packets, in the form of transport stream packets, then a CPCS trailer would add unnecessary overhead. Several compression algorithms may use the MAAL. Therefore, we propose to leave the packet segmentation function to upper layers. The advantage of this approach is twofold: firstly, if the segmentation is performed by the application it will be optimized reducing the overhead to the minimum required. Secondly, the CPCS will not perform any function, thus reducing the processing delay. This, however, requires from the application or an intermediate protocol layer to know the requirements of the MAAL to perform the segmentation according to the MAAL delineation of 47 octets. Such a proposal goes in the sense of the genericity and low delay design principles formulated at the beginning of this chapter.

Passing corrupted SDUs to the upper layers is inherent to multimedia data. Since multimedia data, and in particular video, tolerates some, yet limited, loss, it makes sense to pass as much data as possible to the application. Two advantages come out of this approach. Firstly, as we have seen in Sec. 3.2.2, the impact that the loss of syntactic information has onto video is much more important than the loss of raw video. If a corrupted packet still has headers on it, the loss impact could be, in general, considerably reduced if the header is sent to the receiver. Secondly, due to the nature of multimedia data, concealment mechanisms such as speech interpolation or early resynchronization, could be used whenever corrupted data is received. In general, the more data is available, the better the concealment mechanisms are. This mechanism may perform even better if used in combination with dummy cell insertion. Some compression algorithms are sensitive to data size integrity due to the utilization of run length coding, especially in video compression. It is therefore useful to insert dummy data when recovery is not possible to keep the data size integrity. Moreover, the use of unallowed codewords to fill the dummy cells could be used as an error indication to the decoder. For example, again due to the utilization of variable length coding, the MPEG-2 standard does not allow codewords with more than 5 zero-bytes. If such a codeword appears, the bitstream is errored. Such a feature is however codec dependent. To not violate the genericity principle we propose the dummy cell mechanism to be *selected by the* *user* or, to be more precise, by the application.

The utilization of open-loop techniques stands out due to the requirements of interactive multimedia applications. Different FEC techniques exist adapted to different types of applications. We have shown that a cell level non-interleaved RSE based FEC mechanism performs several orders of magnitude better than an equivalent packet level protection mechanism. The combination of a cell level sequence numbering with the cell-oriented FEC achieves a very reliable transmission.

To alleviate the overhead generated by such a FEC mechanism, we take advantage of the fact that in multimedia not all data has the same importance as explained in Sec. 3.2.2. Since the loss of syntactic information has a larger impact than the loss of semantic data it could be advantageous to selectively protect the *perceptually relevant* information. This however, requires an a priori knowledge of the structure of the data to be transmitted. To keep the AAL generic, we propose the selection mechanism to be driven by higher layers. In other words, the AAL will generate the redundancy data upon request of the upper layer in a cell basis. The algorithm that will identify the perceptually relevant information is developed in Chap. 6.

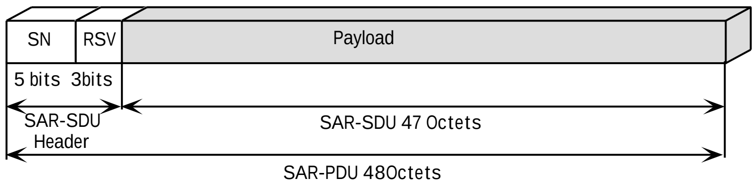


Figure 4.16: *MAAL structure.*

Using a selective FEC mechanism means that the total number of cells per packet will vary. We have proposed to transmit constant packet sizes to reduce overhead and simplify the AAL. However, to perform data recovery and improve reliability we propose to transmit the number of FEC cells per packet to the receiver. This information will be included in the RSV field of the SAR-SDU header. In addition it is possible to add a fourth cell type to be carried by the PTI field to identify FEC cells. We assume in this case that Single Segment Messages (SSM), messages composed by a single cell will never be transmitted.

In summary, the structure of the proposed MAAL, depicted in Fig. 4.16 will be as follows:

- SAR-SDU size of 47 octets
- SAR-SDU header of one octet divided into two fields: a sequence number field and a reserved (RSV) field that will include at least FEC number field
- selective SAR-level FEC
- null SSCS and CPCS.

Such an AAL follows our three design principles. It is generic since there is no application specific function included, even if some functions are user configurable. It is reliable because it provides an efficient cell loss detection mechanism in addition to a cell level FEC scheme. Finally, by deliberately limiting the AAL functions, we keep processing delays to small values fulfilling the third design principle, low delay.

Chapter 5

Multimedia AAL Performance and Specification

5.1 Introduction

This Chapter is devoted to the test and specification of the MAAL according to the proposal of Chap. 4. The proposed functions have been modeled and implemented in a simulated environment. Performances are evaluated from the network point of view using classic metrics such as cell and packet loss as well as from the user viewpoint using perceptual metrics.

We first study the behaviour of the cell loss process by means of spectral analysis in order to validate the approach of Chap. 4 and to evaluate the efficiency to be expected from the proposed cell loss correction method. The next step consists of evaluating the performance of the MAAL from the network point of view. We discussed in Sec. 3.2.2 the difficulties to map network QoS parameters to user QoS parameters for multimedia applications and especially for video information. In order to present a complete performance evaluation that is also meaningful to the end-user, we make use of the NVFM metric to map the obtained network performance results to a user level. We evaluate the impact that cell loss has onto video and also the improvements achieved with the MAAL. The emphasis put onto video is due to several reasons. Firstly, video data will be the main consumer of the bandwidth required by a multimedia application. In comparison audio which is generally encoded as CBR requires little bandwidth. Secondly, some compression algorithms such as MPEG-1 and 2 include audio compression. The audio bitstream is multiplexed with the video resulting in a single output stream. Therefore, from a networking perspective, the results also apply to audio flows.

To show the interest of the proposed approach, all performance results are compared to an equivalent implementation for the transmission of MPEG-2 video applications, based on AAL5, as specified by the current ITU-T and ATM Forum standards [127, 155] (see Secs. 3.5.1 and 3.5.2 for details).

5.2 Simulation Framework

Studying performance of network elements could generally be made in three ways; the analytical which tries to derive closed formulas to evaluate performance parameters, the experimental which uses real systems and the simulated.

The performance studies carried out in this thesis have been performed in a simulated environment. Multiple reasons justify this approach. Simulated environments have the advantage of being totally controlled. The background traffic, the measurement points, the replication of the simulations, the possibility to fix the initial conditions are major advantages of such an environment. Also, a simulated environment allows to change network topologies, sizes and characteristics, feature that real systems hardly allow to do.

Simulation, however, has also drawbacks. Assumptions are necessary to model the network elements which automatically drives inaccuracies. It is much easy to obtain wrong results due to inappropriate modelling, in particular, of traffic sources. This is, however, a problem also found in analytical approaches. Last but not least, the time required to collect data with enough confidence is by far more important than for real systems and can even make simulations impossible to perform.

Even though, we have considered that the flexibility and control inherent to simulation was a major need to perform extensive testing of the proposed protocol mechanisms.

5.2.1 The OPNET Simulation Environment

The simulation framework developed for our purposes is based on a comercial package. The OPTimized Network Engineering Tool -OPNET- [164] is an environment which allows for the development of hierarchical object-oriented network models. The hierarchical structure of this environment allows to develop a network model at different levels as shown in Fig. 5.1. In the network level the topology of the network is defined by interconnecting nodes via links. Each of the nodes contains as many objects as required. Each object could be a network element or a group of elements. The third and lower level allows to specify in C code the functions that each of the objects has to perform. This hierarchical organization allows to reuse and modify in a fast and easy way any network element at any level without modifying the general structure of the simulator.

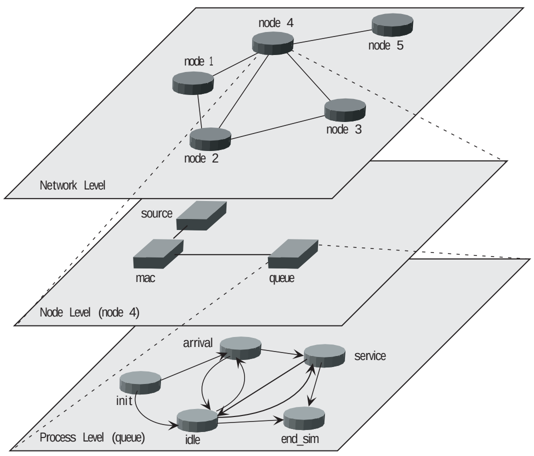


Figure 5.1: *The OPNET simulator environment.*

The OPNET simulator is event-driven, that is, the time between events is skipped. The simulator relies on an *event list* containing the time occurrence of future events. This technique speeds-up simulation executions compared to time-driven simulators.

The different simulators developed are based on the following network elements:

- VBR Video Source

- Bernoulli Source
- Network Adaptation Layer
- Multimedia AAL
- AAL5
- ATM Switch.

The OPNET environment is available in several UNIX workstations. The executable code allows to launch simulations from the UNIX shell via scripts. The scripts are used to set the simulation parameters and to create simulation campaigns with different initial conditions.

5.2.2 Network Setup

One of the problems related to performance evaluation is the simulation scenario. It is obviously impossible to embrace all the situations that may happen in a network. It is therefore necessary to find a scenario general enough to cover as many situations as possible. This requires careful modelling of the network elements, in our case, the ATM switches, the traffic sources as well as the network topologies.

The proposed simulation scenario is depicted in Fig. 5.2. The simulator is composed of four multimedia workstations and two ATM switches. The workstations are connected as two point-to-point communications. Both switching stages, implemented as multiplexers with limited buffer size, are loaded with background traffic provided by several sources. Since our work covers principally real-time services, the buffer sizes have been set to 100 cells. This leads to a maximum queueing delay of 12.47 msec for a 34 Mbps link and 274 μsec for a 155 Mbps link.

The switching stages have been implemented as FIFO multiplexers. This model has been widely used in performance evaluation and has been proven to accurately model a general switch behavior.

To be as general as possible in our experiments, we have used two types of background sources providing both uncorrelated and correlated traffic profiles. To generate uncorrelated traffic we have used Bernoulli sources which are widely used to simulate a multiplex of traffic such as the one that could be found at the entrance of an ATM switch.

By using such a background traffic profile, we are assuming that the background traffic comes from a *large number of independent sources.*

To generate correlated background traffic we have opted for the utilization of real VBR video traces. An advantage of using these traces is that they are not models so there is no accuracy problem in the sources themselves. Several video sequences have been encoded in MPEG-2 as described in the next section. These video sequences have different characteristics in terms of scene changes and motion which lead to different traffic profiles covering a relatively large set of correlated VBR sources. The characteristics of the different video sequences used are described in Sec. 5.2.3.

The multimedia workstations, which generate the traffic under test (TUT) implement either AAL5 or the MAAL. We have tested both CBR and VBR traffic using real video sequences as described in the next section.

To guarantee the same CLRs to both cell streams, the background traffic is replicated and sent simultaneously to both multiplexing stages. Since all background video streams have been encoded with the same GOP size and quantization factor (see Sec. 5.2.3 for details),

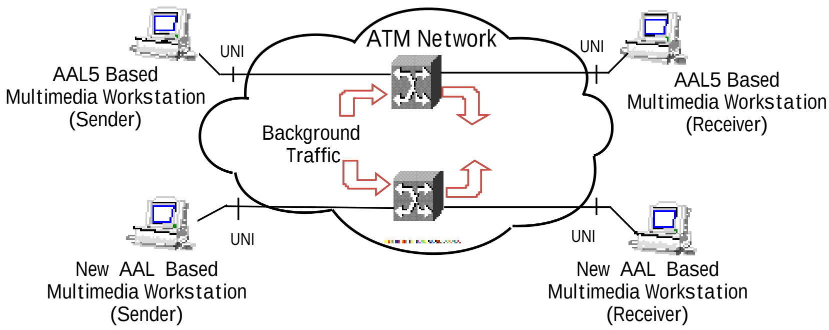


Figure 5.2: *Simulation scenario.*

correlations between different sources may appear. To avoid such correlations random phases are used at simulation startup. The phases have been calculated on a PDU basis instead of on a frame basis to reduce the probability of simultaneous frame transmissions. Also, to avoid transient phenomena, the background sources begin transmission before the TUT.

The TUT is sent to the switch where it is multiplexed with the background traffic. Since the switch buffer is limited in size, some of the TUT or background cells may be lost. The TUT is then routed to the receiver end system where the data is reassembled. The background traffic is assumed not to interfere further with the TUT and thus is directly routed to a traffic sink after leaving the switching stage.

5.2.3 Video Setup

The foreground traffic used for our experiments has been obtained by encoding real video sequences. By doing so we avoid the accuracy problems due to mathematical models of VBR video sources. In addition, to be able to use a video quality metric, we needed real bit streams. We have performed the measurements with MPEG-2 streams for several reasons. Firstly, the utilization of predictive algorithms worsens the impact of data loss by adding a temporal propagation component which is much more annoying than a spatial one only as discussed in Chap. 3. Secondly, because the major compression algorithms namely, H.261, H.263, MPEG-1 and MPEG-2 use such predictive techniques, the results achieved with MPEG-2 could be interpreted without loss of generality. Thirdly, because the foreseen future MPEG-4 standard will at least encompass MPEG-2, the results will still be valid. Last but not least, because we believe that MPEG-2 will be increasingly used for all types of audiovisual applications.

The main video sequence used to generate the traffic under test (TUT) consists of a ski sequence of 1000 frames of ITU-R 601 format (720×576). We have generated both CBR and VBR video streams with a Test Model 5 (TM5) MPEG-2 software encoder [165] as interlaced video, with a structure of 12 images per GOP, 2 B pictures every reference picture and a single slice per line (i.e. 45 macroblocks per slice).

To generate the VBR video, we have used the encoder in open-loop rate control with a constant quantizer scale. The sequence provides several scene changes as well as fast motion. Figures 5.3 (a) and 5.3 (b) show two frames of two different scenes of the sequence.

All video processing has been performed offline. The data transmitted in the simulations are files containing the actual video data. Therefore to generate VBR video we have used a timebase file containing the transmission time of every TS packet. Since the software encoder does not provide any timing information we have created the timebase by calculating the

(a) Sample 1 (b) Sample 2

Figure 5.3: *Samples of the original video sequence used for transmission.*

	Peak Rate (Mbps)	Mean Rate (Mbps)	Peak-toMean Ratio
baseball	9.722	3.2	3.04
Mobile	12.445	5.34	2.33
basket	13.275	9.256	1.434
commercial	7.76	3.43	2.26
flower	12.238	7.32	1.67
ski	7.36	4.19	1.75
football	10.722	4.575	2.34
Barcelona	15.838	5.3	2.98
Movie	9.253	5.672	1.63

Table 5.1: *Main traffic parameters of the background video streams encoded with a quantization factor of 26.*

time required to transmit all the TS packets of a frame in a timeframe slot (i.e. $\frac{1}{25}$ sec or $\frac{1}{30}$ sec. The timebase file, which gives a TS granularity, allows to calculate the interarrival times of a PDU independently of the number of TS packets composing the PDU. Such technique automatically generates a *shaping* of the TS packets since the interarrival times calculated are uniformly distributed across a timeframe slot.

The video sequences used as background traffic have the same characteristics as the TUT sequence and have been generated using the same processing described above. A single quantization factor of 26 has been used for these sequences which gives a CATV video quality. Tbl 5.1 gives the main characteristics of the video streams.

The MPEG-2 decoder provides some error concealment techniques. These techniques have been used for different reasons. The first is to be consistent with real implementations and the second one to be able to perform the perceptual measurements. Indeed, the human visual system models currently developed and the derived metrics have been tested for errors below the *suprathreshold* defined in Sec. 3.2.2. The problem is that in general, the errors due to cell losses generate highly visible artifacts in the sequence and these errors are all above this suprathreshold. By using concealment techniques, the artifacts could be considered as being below, making the video quality metric accurate.

5.2.4 Performance Metrics

When performing simulations it is necessary to define the metrics to be used to evaluate the performance of the tested system. As described in Sec. 2.2.4 the traffic performance parameters defined in an ATM traffic contract are the Cell Loss Ratio (CLR), the Cell Delay Variation (CDV) and the Cell Transfer Delay (CTD).

CDV and CTD are generally splitted into two major components: the terminal equipment component and the network component. The former includes all delay and delay jitter generated by the processing and synchronization of the end-user equipment and the access system. Therefore this components partly depend on the AAL design. The MAAL has been developed to achieve low delay and jitter. The latter component can be further decomposed into a transmission delay and a queueing delay. This part is totally *independent* of the protocol layers used and depends only on the network topology and *traffic conditions*. We therefore do not consider CDV and CTD as the main performance parameters for the AAL.

This thesis mainly focuses on the cell loss aspects and the recovery functions which depend on the AAL design. We therefore have taken the CLR as one main performance parameter.

While the CLR, as defined in recommendation I.356 [40], is related to the ATM layer performance, we have focused on the CLR as observed by the end-user. That is, we measure the CLR after reassembly of the packets. Providing such a measure allows us to characterize the CLR experienced by an end-user which uses an AAL with a packet discard mechanism. Such an approach is justified by the fact that we do not consider a retransmission based mechanism. In such a case, the actual CLR is therefore fixed by the reassembly strategy of the used AAL. As it is shown in the remainder of this chapter the reassembly mechanism of the AAL leads to very different CLR values than those measured at the ATM layer. This heavily influences the quality of multimedia data and in particular of the displayed video information.

When performing measurements it is important to know the degree of confidence one can have on the observed values. This is particularly important in simulations where the number of samples is not always enough to guarantee accurate results. The CLR metric is defined as the mean number of cells lost neasured during a given period of time. However, such a metric has little significance if the confidence interval is not known. A cell loss process L_n is a $0/1$ time-series process of duration n. If we define the random variable L associated to the loss event we define:

$$L = \begin{cases} 1 & if \quad loss \\ 0 & if \quad no\ loss \end{cases}$$

Then the loss probability p is given by the estimation of the mean of L. If L_n are n independent trials and q denotes the number of successes (losses) then an estimate of the loss probability p is given by:

$$\hat{p} = \mu_l(n) = \frac{1}{n}\sum_{i=1}^{n} l(i) = \frac{q}{n}.$$

Due to the central limit theorem, $\hat{p}$ can be approximated by a normal distribution with mean $E[L_n]$ and variance $\frac{Var[L_n]}{n}$ for large values of n. For a Bernoulli process, the sample mean is given by $\hat{p}$ and the sample variance by $\hat{p}(1-\hat{p})$. In this case, the $100(1-\alpha)$ percent confidence interval for $\hat{p}$ is given by:

$$\hat{p} \pm \left(z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right), \tag{5.1}$$

where z_β is the β quantile of the standard normal density.

When the process appears to be correlated Eq. 5.1 does not apply and it is necessary to obtain n independent samples of $\hat{p}$. This is achieved by performing *n independent runs* of the same simulation. Then the $100(1-\alpha)$ percent confidence interval for p is given by:

$$\mu(n) \pm \left(t_{n-1,1-\alpha/2} \sqrt{\frac{\sigma^2(n)}{n}} \right), \tag{5.2}$$

where $\mu(n)$ is the average and $\sigma^2(n)$ is the variance of the n estimations and t_{n-1} is the Student distribution with $n-1$ degrees of freedom.

The second performance metric used in our tests is the Moving Pictures Quality Metric (MPQM) and its improved version Normalized Video Fidelity Metric (NVFM) described in Sec. 3.2.2. The MPQM metric generates a time-series Q_n where $q(n)$, a particular realization of Q_n, is the perceptual quality measured for frame number n as shown in Fig. 5.4. The final quality evaluation is obtained by averaging Q_n for a given CLR. Since the values of the time-series are not independent, confidence intervals are given by Eq. 5.2. Multiple runs are therefore necessary to obtain reliable estimates of $E[q(n)]$.

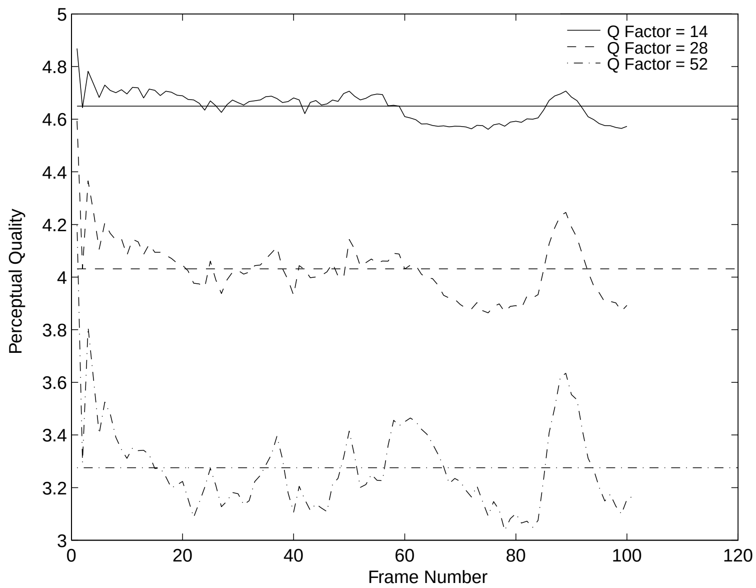


Figure 5.4: Example of perceptual quality time series.

5.3 Characterization of the Cell Loss Process

The design of the MAAL derived in Chap. 4 assumes a non-correlated cell loss process according to the low traffic source model. Key values such as the sequence number field size of the MAAL SAR-PDU and features such as the cell level error recovery find their rationale in the assumption of iid cell loss processes. This hypothesis has proven to be valid for independent periodic sources, such as CBR. No proof, however, has been given concerning the validity of this approach for more general VBR video traffic models. Therefore the first task to validate our approach is to characterize the cell loss process for the proposed configuration.

The metrics we have used to characterize the cell loss process are the *Power Spectral Density* (PSD) and the autocorrelation functions of the L_n process. The PSD is defined as the Fourier transform $F$of the autocorrelation function $R_l(n)$ for all lags as follows:

$$\Phi_l(f) = F\{R_l(\tau)\} = \int_{-\infty}^{\infty} R_l(\tau) exp(-j2\pi f\tau) d\tau.$$

For our purpose we use the autocovariance function of L_n, that is the autocorrelation function of the zero-mean process $L_n - \mu_l$. Then the PSD is expressed as follows:

$$\Phi_l(f) = F\{C_l(\tau)\} + \mu_l^2 \delta_f, \tag{5.3}$$

where $C_l(\tau)$ is the covariance function μ_l, the mean value of the loss process and δ_f the impulse function.

Our simulation study is performed in two phases. We first study the cell loss process for CBR transmission. Even if this is not necessarily a realistic scenario since CBR communications will use deterministic allocation, the study of CBR is used as a first step to validate the metrics as well as the simulator of Fig. 5.2 by comparing the results to those found in [57]. In addition, other multimedia sources such as audio produce CBR streams but with much lower bit rates. The second phase focuses on the cell loss process of VBR connections under different types of background traffic.

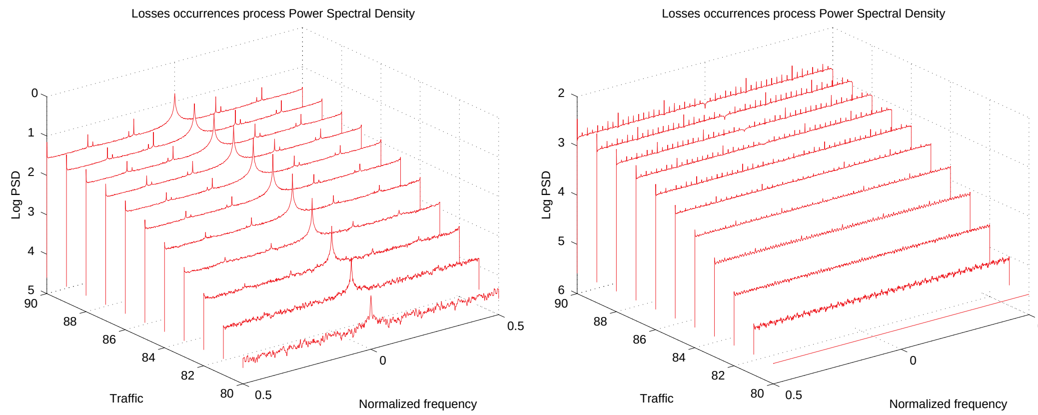
5.3.1 Preliminary Study: Cell Loss Process for Constant Bit Rate Traffic

According to [57], the cell loss process of a periodic source can be modelled by an iid process under the condition that the fraction of the link rate used by the observed connection is small enough (around 10%) and the number of background sources is large. Note that the buffer size is said *not to have any influence* on the cell loss process. We study in this first set of simulations, which uses Bernoulli background traffic, the influence of the background load on the cell loss process.

We have used a TUT load of 16% which is slightly superior to the value found in the literature. This value may seem to be excessive considering that standard link speeds are of 155 Mbps and beyond. However, we use this value because we want to include in our studies the cases in which links of 34 or 45 Mbps are used. This leads to application bit rates of 5.5 and 7.2 Mbps respectively which correspond to medium to standard TV quality. The generalization of the results obtained with these link rates are easily generalized to higher capacity links.

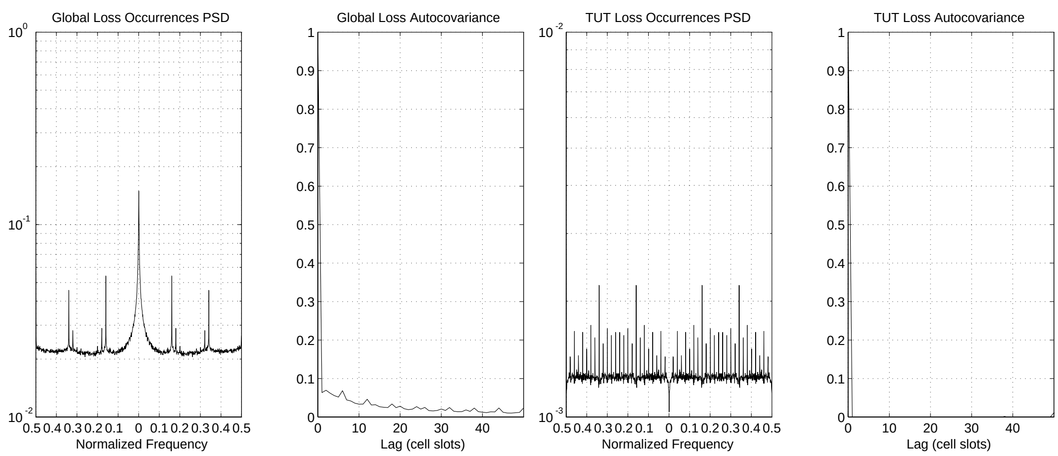
The background traffic load varies to obtain different CLR values. Note that in some simulations, the overall offered load is higher than 100%. Bearing in mind that we want to test the behavior of the cell loss process we have set high network utilizations to achieve high CLRs and generate enough samples to achieve reliable measurements.

Figure 5.5 depicts the PSDs of both the global loss process $l_g(k)$ which embraces all connections together, and the TUT loss process, $l_{TUT}(k)$, as functions of the background traffic load. The spectra of the $l_g(k)$ processes exhibit an important lobe at low frequencies which tends to be flat at higher frequencies. Also, the amplitude of $\Phi_g(0)$ indicates large fluctutations around the mean value. This behavior suggests a correlated loss process. This is to be expected because during congestion periods, consecutive cells are dropped by the switches. Figure 5.6 (a) partly confirms this correlated behavior. The autocovariance shows that a very low correlation, less than 0.1, exists at least up to 50 cell slots. This means that $l_g(k)$ is not very bursty, since the congestion periods are of short duration. 50 cell slots at 34 or 45 Mbps correspond to 12,4 and 9,4 μsec respectively. Therefore there is a small probability of occurrence of consecutive losses. Note also that when the traffic intensity decreases the peak also decreases and some noise appears in the curves. This is due to the relatively low frequency of the loss occurrences.



(a) Global Process (b) TUT Process

Figure 5.5: *Cell loss process power spectral density. TUT load 16%.*



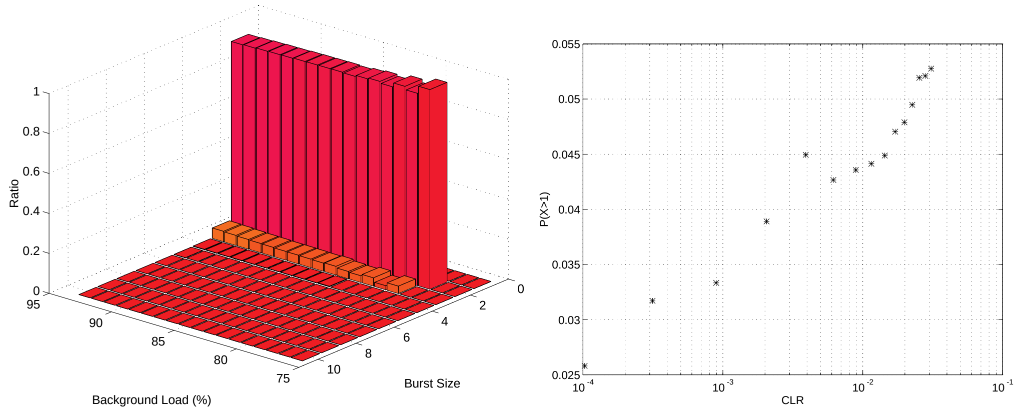
(a) Global Process (b) TUT Process

Figure 5.6: *Power spectral density and lag plot for CBR TUT traffic.TUT load 16%.*

Comparatively, the PSDs of a single CBR connection (Fig. 5.5 (b)), that is, the TUT, shows no correlation. The spectra appear flat except for the peaks observed at regular intervals. Again, the lag plot of Fig. 5.6 (b) confirms this observation. These peaks which are visible in both PSD plots suggest that a periodic component exists. Note that the intervals remain constant for all CLRs.

If we assume that a periodicity exists in the sampled processes then the peaks have to appear at multiples of the period of the signal, namely, $\frac{1}{T}, \frac{2}{T}$, etc.... We can then calculate the period of the process. The first peak, more visible in Fig. 5.6, appears at a frequency of 0.16, therefore the period is 6.25. This indicates that this periodic components is due to the fact that cell losses occur at precise instants since the cells are sent at a constant rate. The TUT load being of 16%, a cell is sent every 6.25 cell slots.

We define the random variable X as being the number of observed consecutive cell losses. Then $P(X > 1)$ expresses the probability of observing two or more consecutive cell losses. The values shown in Fig. 5.7 (b), which displays $P(X > 1)$ as a function of the CLR, are relatively low, below 6.10^{-2} for the worst case CLR. In addition, the histogram of Fig. 5.7 (a) which depicts the burstiness of the cell loss process indicates that when this occurs, the number of consecutive losses seldom exceeds two cells. Therefore, cells are seldom lost in clusters. No evidence of burstiness appears from these figures.



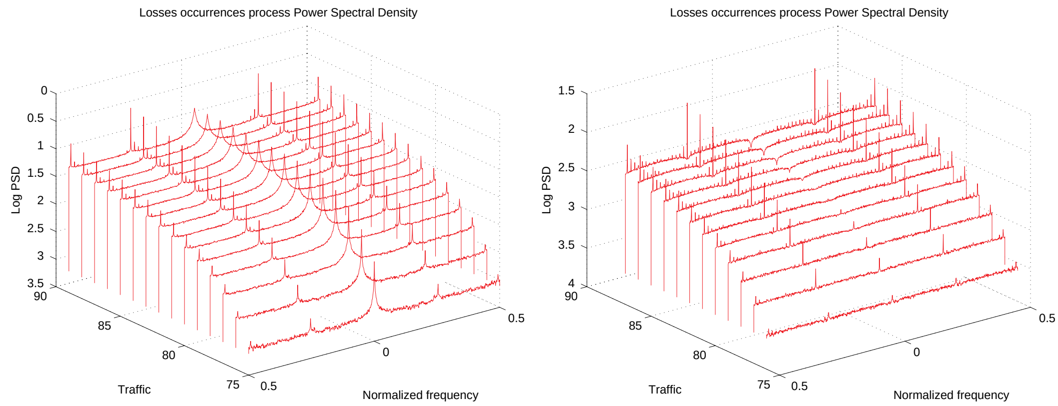
(a) $P(X = k)$ (b) $P(X > 1)$

Figure 5.7: *Loss occurrences process characterization of a CBR stream. TUT load 16%.*

Given this first set of results we now change the TUT load to 25% of the link load. This leads to an application bit rate of 8.5 and 11,25 Mbps respectively if we again consider 34 and 45 Mbps links. If using values of 25% of the link rate proves to still fulfill the iid model then we can consider that for a range of bit rates able to cover a wide range of multimedia applications, given that the rest of the traffic comes from a large number of connections, the cell loss process would be accurately approximated by the iid model.

The figures obtained using 25% of TUT load show a behavior close to the precedent case. The global loss process still exhibits an important lobe at low frquencies while the TUT loss process still appears almost not correlated. The major difference compared to the precedent case is that the periodic component is more important. This is due to the fact that the fraction of CBR cells lost is higher than for the precedent case which leads to larger amplitude peaks. Here again we verify that the peaks are due to the CBR traffic. The TUT load being of 25%, a cell is sent every 4 cell slots. Therefore the period of the process should be 0.25 which is the frequency obtained in Figs. 5.8 and 5.9.

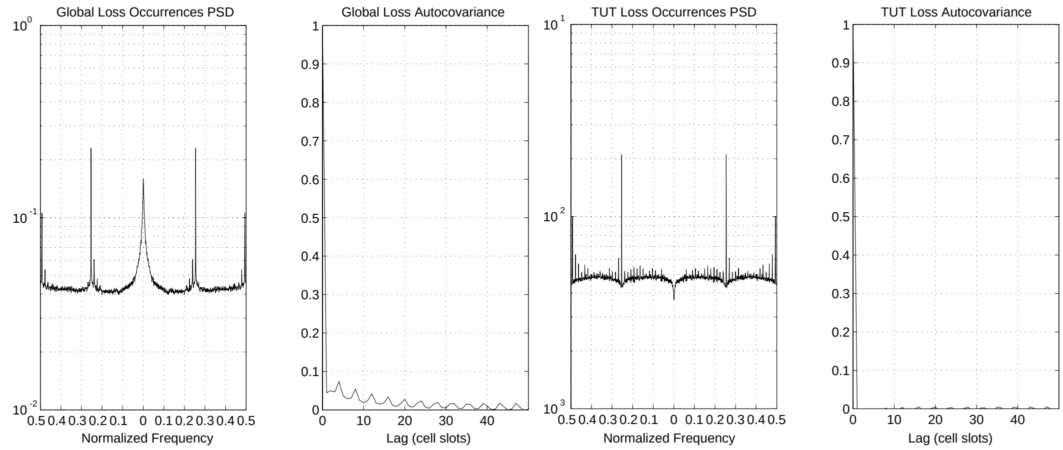
These observations confirm the results already presented in the literature which state that the cell loss process of a single CBR connection can be modelled by an iid process if the fraction of the link rate used by the observed connection is small and if the background traffic is provided by a large number of sources which leads to a Bernoulli traffic model. The figures show that for TUT traffic values of up to 16% no correlation is visible in the cell loss



(a) Global Process

(b) TUT Process

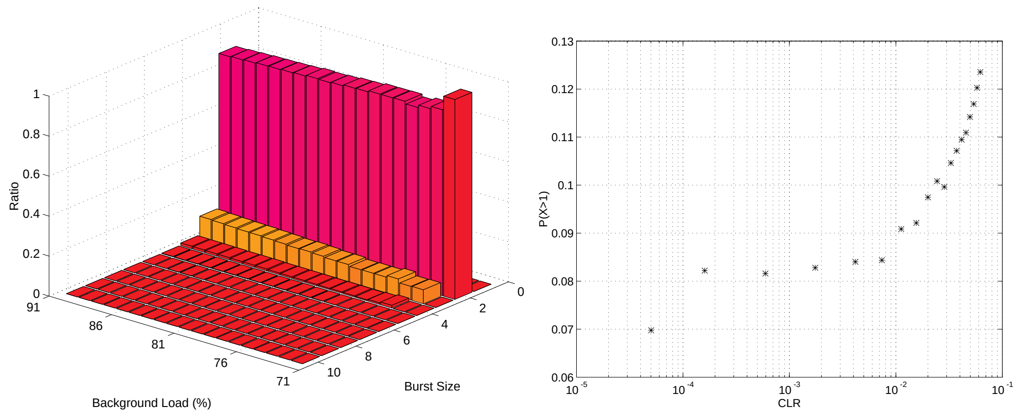
Figure 5.8: *Power spectral density of the cell loss process of a CBR stream. TUT load 25%.*



(a) Global Process

(b) TUT Process

Figure 5.9: *Power spectral density and lag plot for CBR TUT traffic. TUT load 25%.*



(a) $P(X = k)$ (b) $P(X > 1)$

Figure 5.10: *Loss occurrences process characterization of a CBR stream. TUT load 25%.*

process of that connection. When the TUT traffic occupies 25% of the available bandwidth, the figures slightly deviate from the iid model. This load value could then be considered as an *upper bound*.

The intuitive explanation of this phenomenon is that due to the cell spacing the probability of observing consecutive cell losses belonging to the same connection is small as depicted in Fig. 5.10 (b). In addition when this occurs the number of consecutive cells lost is limited to small values which in our tests reaches 4 to 6 for very high CLRs as described in Tbl. 5.2. In other words during the fraction of time that a congestion period occurs, the incoming and probably lost cells belong to a *large number of different connections*, thus reducing the probability of losing a large number of consecutive cells from the same connection. This behavior is also explained by the global cell loss process which exhibits a very short term correlation with almost no burstiness. Since the majority of the traffic comes from the Bernoulli sources, this traffic prevails and partly hides the effects due to the CBR traffic.

TUT Load (%)	16	25
Max Burst Loss Size	4	6
CLR	3.4e-2	6.9e-2

Table 5.2: *Maximum loss values for CBR traffic.*

Considering the precedent observations, we conclude that for the proposed scenario, from the point of view of a single connection, the loss process is *almost not correlated* and can be approximated by an iid process. This matches the results found in the literature validating our simulation environment and the MAAL desing approach considered in Chap. 4.

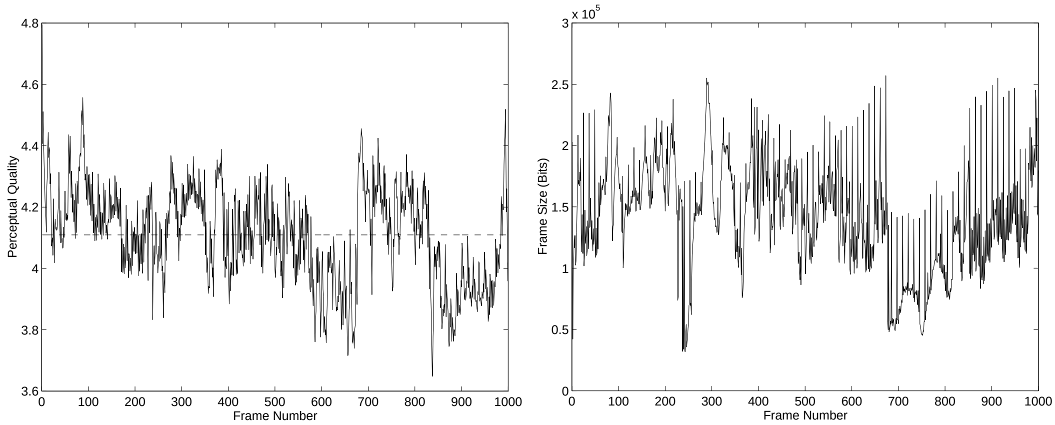
5.3.2 Cell Losses for Variable Bit Rate Traffic

The second phase of the study, consists on replacing the foreground CBR source by a VBR encoded video stream. Due to the nature of open-loop VBR compressed video, which has been proven to be self-similar [56], we can expect the loss process to present some burstiness. During high activity periods where high bit rates are generated, the probability of observing consecutive losses from the same source should increase leading to some correlations on the cell loss process. However, as shown in the precedent section if the percentage of Bernoulli traffic is high enough then the cell loss process tends to be non-correlated because the contribution of the TUT to the loss process is not strong enough.

To obtain a representative VBR stream we have generated from the same original sequence three compressed video streams. To cover as many traffic characteristics as possible in terms of burstiness factors, defined here as the *peak-to-mean* ratio, as well as peak and mean bit rates we have used three quantization factors leading to different video qualities. The first considered, 52 gives an image quality of 3.27 which achieves a fair quality according to the classification of Tbl. 3.2.2. The second video stream uses a quantizer scale of 28 which gives a perceptual quality of 4 close to CATV video quality. The third quantizer scale, 14, achieves the highest encoding quality. Note that as illustrated by Fig. 5.11 an open-loop encoder does not output a constant quality stream. However, the better the encoding quality, the smaller the variations around the mean value. The main traffic and perceptual characteristics of the video streams summarized in Tbl. 5.3 show that the burstiness decreases while increasing the encoding video quality. This is because the average bit rate increase is more important than the peak bit rate increase measured over a *large period of time*. Fig. 5.11 (b) shows the frame size trace of the compressed sequence encoded with a quantization factor of 28.

Quantizer	14	28	52
Burstiness	1.68	1.80	2.06
Peak Bit Rate (Mbit/s)	11.53	6.56	4.65
Mean Bit Rate (Mbit/s)	6.84	3.64	2.25
Average Perceptual Quality	4.65	4	3.27

Table 5.3: *Main traffic and perceptual parameters of the encoded video streams.*

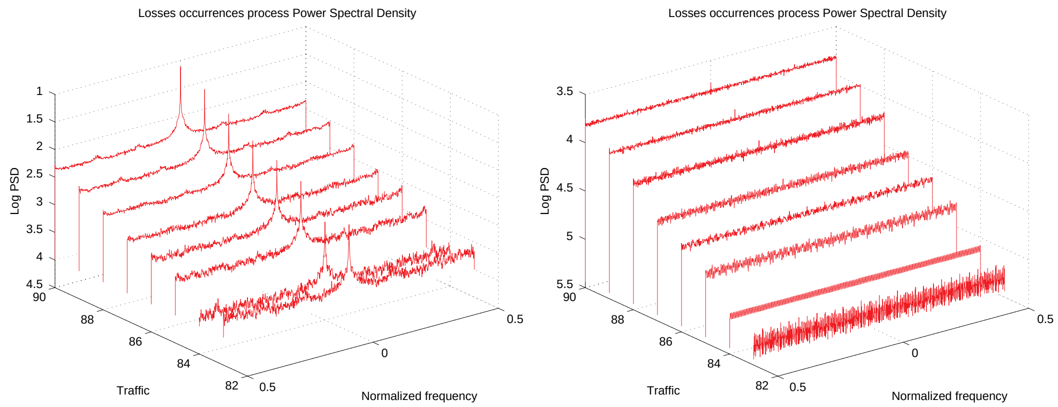


(a) Perceptual Quality (b) Bits per Frame

Figure 5.11: *Perceptual quality and size per frame for the ski sequence encoded open-loop. Quantization factor 28.*

5.3.2.1 Influence of Traffic Load and Burstiness

In general, VBR does not achieve high *average* link bandwidth utilizations as can be seen from Tbl. 5.3 if 155 Mbps links and higher are considered. For 34 and 45 Mbps links, it is still possible to have relatively large mean link utilizations but for very high quality compressed video. It is therefore possible that in peak activity periods, the percentage of link bandwidth used reaches high values. In these cases, cell losses may occur in clusters leading to some correlations. What will be determining is the frequency of these clustered losses. If it is high, then correlations will appear. In the opposite case they will not be visible because the average number of clustered losses will be small compared to the single cell losses and therefore will appear as uncorrelated.



(a) Global Process

(b) TuT Process

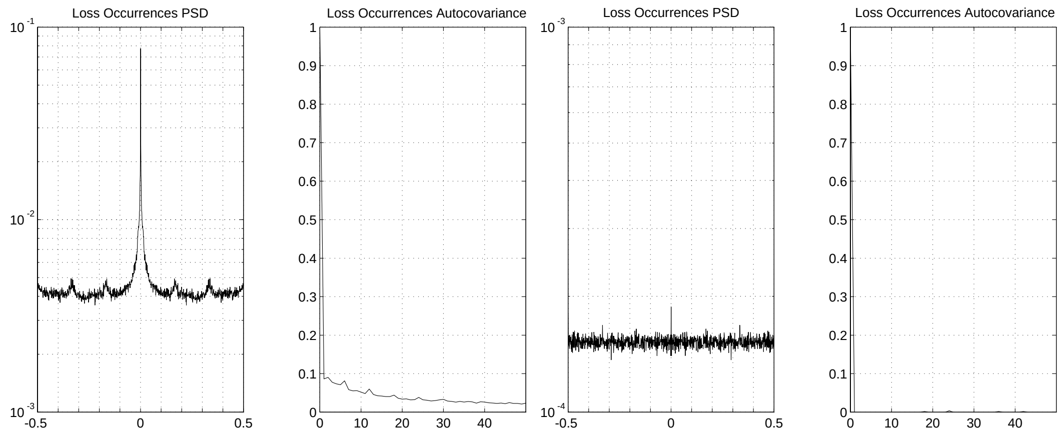
Figure 5.12: *Power spectral density for VBR encoded video: Quantization factor 52*

Figure 5.12 shows the PSD functions of $l_g(k)$ and $l_{TUT}(k)$ for the video stream encoded with a quantizer of 52. The PSD functions of the global cell loss processes show, like for the CBR case, a very low correlation. The autocovariance function of Fig. 5.13 (a) shows values slightly higher than for the CBR case (see Fig. 5.6 (a)) but still below 0.1 for all lags.

Concerning the TUT spectra, a visible peak appears at $\Phi(0)$ for all loads which according to Eq. 5.3 corresponds to the mean value of the process. However, for the rest of the frequencies all spectra appear flat denoting no correlation. As shown in Tbl. 5.3 the peak cell rate of the video stream compressed with the highest quantizer scale (i.e. the lowest quality) is relatively low (around 4.65 Mbit/s) and the average link utilization is less than 6%. The consequence of this is that since the fraction of the link rate used is small even at peak rate, the cells are spaced in such a way that correlations almost do not appear in case of loss as illustrated by the autocovariance of Fig. 5.13 (b).

If we compare the PSDs of the video streams encoded with the quantization factor of 52 to the CBR cases we see that they both look very close. The spectrum is flat denoting no correlations. This is mainly due to the low bit rate and very low link utilization. The main difference between both spectra is that the VBR case does not have the periodic component observed for the CBR case which is consistent with the proposed explanation.

The second TUT stream corresponding to a CATV video quality shows almost the same behavior as the precedent case. The spectra of $l_{TUT}(k)$ appear to be flat except for the



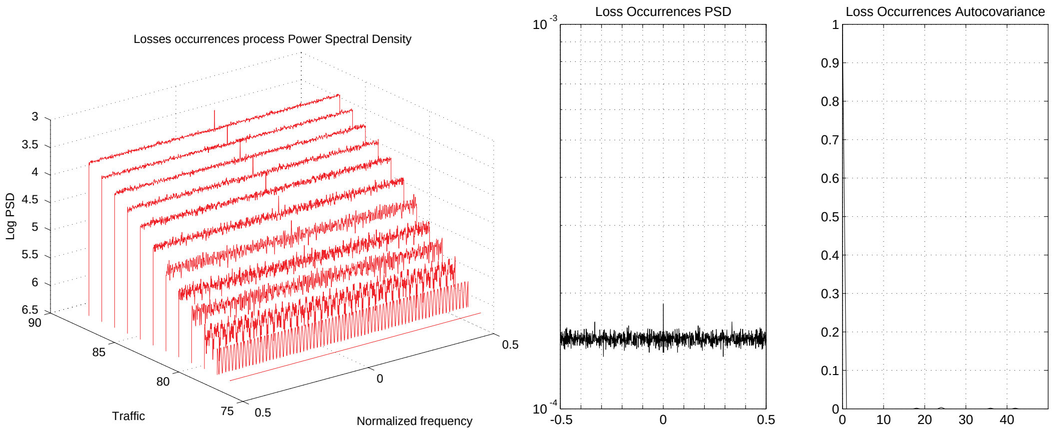
(a) Global Process (b) TUT Process

Figure 5.13: *Power spectral density and lag plot for VBR encoded video: Quantization factor 52. Background load 90%.*

impulse at the origin (Fig. 5.14). However, the lag plot that appeared flat for the precedent case shows some non-zero values for lags between 10 and 40. Albeit negligibles these values may indicate that the increasing bandwidth required by the VBR stream, and in particular the higher peak rate (up to 19%), begins to generate some consecutive losses. The PSDs for the global process, not shown here present the same characteristics observed in the CBR experiments.

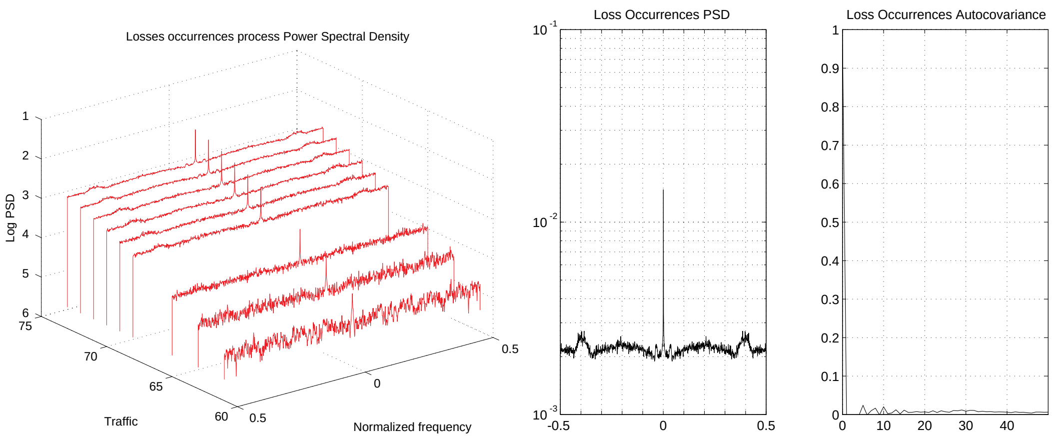
Finally the high quality video stream corresponding to a quantization factor of 14 seems to confirm the trend observed in the precedent experiments. The global loss process, not depicted in the figures, shows the same low correlation values as in the precedent case due, mainly, to the influence of the Bernoulli traffic. However, the TUT PSDs do not look flat. The spectra depicted in Fig. 5.15 (b) shows that visible lobes, becoming more apparent with increasing CLR and load, appear at high frequencies. The autocovariance plot shows however, that even if some non-zero values appear, they still may be considered as negligibles.

Unlike for the CBR and low quality video cases where the cells are spaced enough to eliminate correlations, the high quality VBR video source produces high PCRs. When this occurs, the probability of consecutive losses is increased because the cells are relatively close in time. As shown in Fig. 5.16 which plots $P(X > 1)$, where X is the number of consecutive losses, this probability increases with the encoded video quality. For the high quality video stream, $P(X > 1)$ is not negligible and varies between 17% and 25%. For the other two cases, $P(X > 1)$ is always below 10%. However, as depicted in Fig. 5.17, when consecutive losses occur they are limited to small values, 2 or 3 cells, even if in very rare occasions and under unacceptably high CLRs these values may be much higher as shown in Tbl. 5.4. The explanation to this observation is that as indicated in Tbl. 5.3, the mean cell rates lead to low fractions of the link rate. However, the PCRs produce peaks which require up to 19% and 30% for the CATV and high quality video streams respectively and cannot be considered as negligibles. In such situations consecutive cell losses are very likely to occur. However, as shown in Fig. 5.11 these high bit rate situations seldom occur and consequently clustered cell losses have a small probability to occur which in turn explains the very small values showed by the autocovariance function.



(a) Power Spectral Density (b) PSD and Lag Plot. Background Load 87%

Figure 5.14: *Power spectral density and lag plot for VBR encoded video: Quantization factor 28.*



(a) Power Spectral Density (b) PSD and Lag Plot. Background Load 75%

Figure 5.15: *Power spectral density and lag plot for VBR encoded video: Quantization factor 14.*

Q Factor	14	28	52
Max Burst Loss Size	10	4	3
CLR	1.35e-1	3.75e-2	1.41e-2

Table 5.4: *Maximum loss values for VBR video streams.*

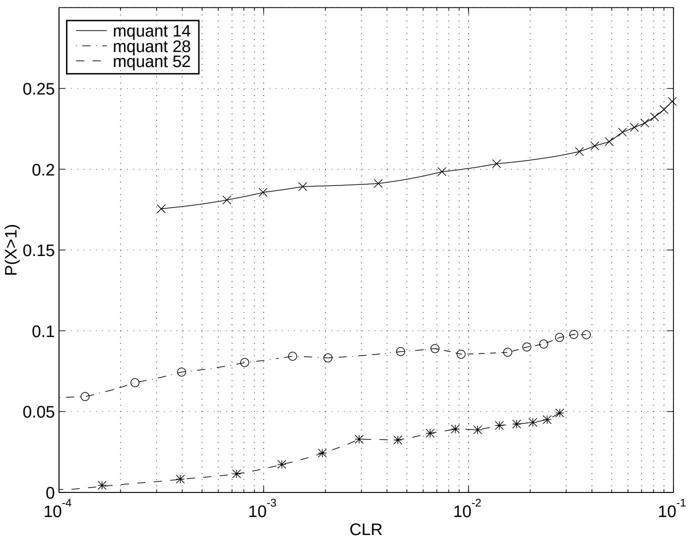


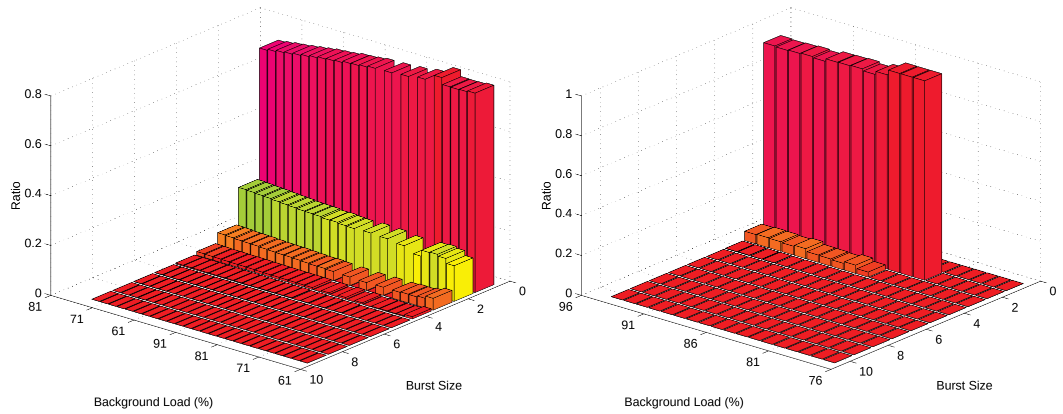
Figure 5.16: *probability of observing two or more consecutive cell losses for VBR traffic as a function of the CLR.*

The fact that the background traffic is of type Bernoulli and therefore not correlated is a major reason to the observed behavior. The probability of observing long periods of congestion due to the bacgkround traffic is very small. The only source which contributes with long bursts of data is the video source. However, since the buffer sizes used are small, the loss process occur at the cell rather than the packet scale.

The results show that while the traffic contribution of the video source *in average* is below 10% (mquant 28) the burstiness is attenuated by the background traffic and therefore no correlations are visible. The last experiment performed with the high quality video stream shows that the video contribution not being negligible anymore, correlations in the cell loss process begin to appear. However, the values of the autocovariance figure still show that the observed correlations are almost negligible.

5.3.2.2 Influence of the Background Traffic Profile

In the precedent studies we have used Bernoulli sources for the background traffic. The reason to apply this model is that we assume that the background traffic is generated by a large number of independent sources. Bernoulli traffic is memoryless and therefore not correlated through time. The results obtained in the precedent paragraphs show that the resulting global loss process is almost not correlated. There are no long congestion periods and the probabilities of observing large bursts of consecutive losses is very small even if a bursty cell stream is multiplexed (see Figs. 5.7, 5.10 and 5.16).



(a) Quantization factor 14 (b) Quantization factor 52

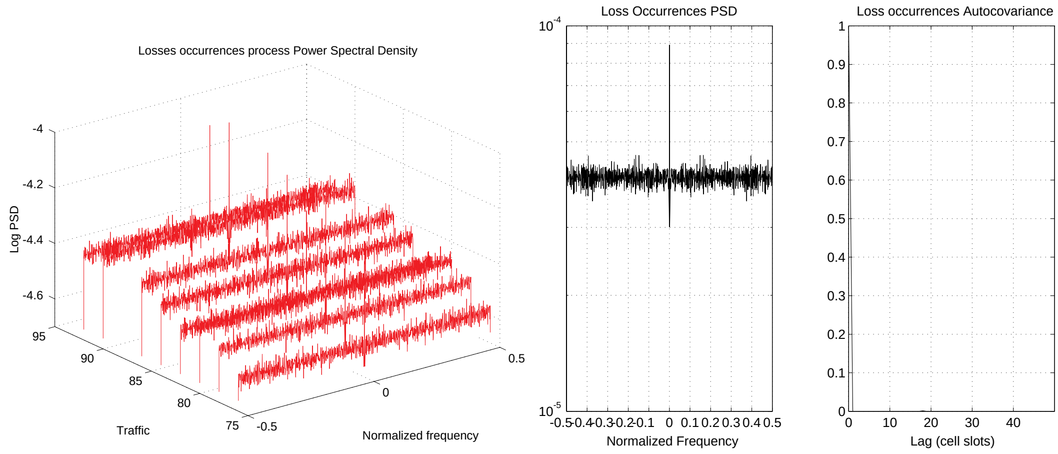
Figure 5.17: *distribution of consecutive cell losses for VBR video traffic.*

This situation may howsoever be different if the background traffic presents some temporal correlation and a bursty characteristic. To study the influence of the background traffic, we have performed a set of simulations with VBR video sources only.

The number of background sources has been set to 15. We achieve an aggregated peak rate of 155 Mbps and a mean bit rate of 75 Mbps. Since the video sources have a fixed rate to change the network utilization we have varied the link bandwidth. To be close to the precedent configurations studied, we have chosen the video sources such as that none of the sources use more than 10% of the bandwidth. The background sources are independent. We have used random phases to avoid as much as possible traffic correlations at the frame or GOP levels. The random phases have been calculated over all the frames in each of the sequences. Since not all the sequences have the same number of frames and the starting point for each of the sequences is uniformly distributed, when a background sequence finishes, it restarts playing from frame 1. Eight different sequences with different characteristics and duration have been used as described in Sec. 5.2.3. They have all been encoded with the same quantization factor of 26.

Figure 5.18 shows the PSD of the TUT process. Albeit relatively noisy, the spectra do not exhibit the lobe characteristic of a correlated process. The peak according to Eq. 5.3 gives the mean value of the cell loss process $l_{TUT}(k)$.

The detail of Fig. 5.18 (b) clearly shows the impulse at $\Phi(0)$ but no lobe is visible. Also, the autocovariance functions shows no correlation of the loss process. These results are consistent with the large deviation theory which states that the multiplex of a large number of independent sources with the same characteristics tends to a normal distribution. Therefore if a large number of video sources are multiplexed the resulting traffic should converge towards a Bernoulli cell stream. Consequently the PSD figures should not exhibit any correlation. This is actually the case. When the background traffic is generated by multiple correlated VBR background sources, the cell loss process still behaves as a uniformly distributed process as long as the sources do not use more than 10% of the link bandwdith leading to a significant potential for *statistical multiplexing gain*.



(a) Power Spectral Density

(b) PSD and Lag Plot. Background Load 85%

Figure 5.18: *Power spectral density and lag plot for VBR TUT traffic under correlated background traffic.*

5.3.3 General Remarks on the Obtained Results

The scenarios proposed for the study of the cell loss occurrences process assumes that the background traffic is generated by a large number of sources, the fraction of the link rate used per source seldom exceeds 10% and that given the nature of the applications the switch buffers are small. Under these conditions we have shown throughout all the experiments that the cell loss process can be accurately approximated by a uniformly distributed process. Precedent studies showed this property for CBR, On-Off and Gaussian sources. We have shown that for *another class of sources*, open-loop VBR video sources which have self-similar characteristics, the approximation is still valid. In addition, we have shown that, concerning CBR sources, for values beyond the 10% limit established the model is still accurate.

The main reason to this accurate approximation is the cell spacing which for CBR reduces the probability of consecutive losses to occur. Concerning VBR traffic whose cell spacing varies, some clustering in the cell loss process has been observed. Since the peak rate achieved by such sources may be several times higher than the average bit rate, in high activity periods the probability of consecutive losses is increased. However, the frequency of occurrence of such clustered losses is small and is actually negligible, thus leading to an uncorrelated loss occurrences process. In addition, the utilization of small buffers tends to favor this situation.

The influence of the background traffic seems to be relative as far as the fraction of the link bandwidth used per source is small, since both Bernoulli and correlated VBR sources show close results.

We have shown the applicability of the low traffic source assumption to a large majority of configurations likely to occur in real networks validating the approach taken in Chap. 4 for the design of the MAAL. The next sections are devoted to study the efficiency of such design related to interactive multimedia communications.

5.4 Multimedia AAL Performance

We have introduced in Chap. 4 a set of functions we consider as mandatory to fulfill the requirements of real-time interactive multimedia applications. The specification of a Multimedia AAL based on this set of functions is partly based on the assumption of a uniformly distributed cell loss process. In the precedent sections we have shown that this approximation is accurate enough to validate the approach used.

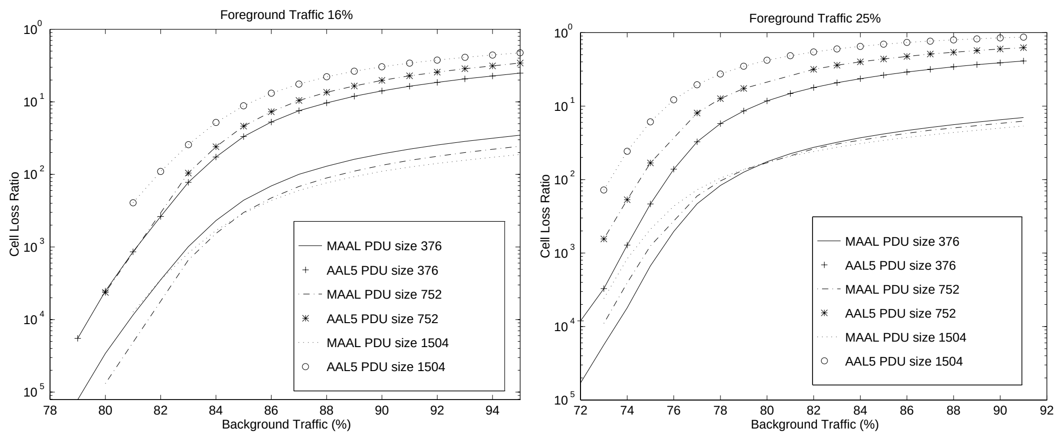
In this section we test the performance of this MAAL by using both network and user oriented metrics introduced in Sec. 5.2.4. Performances are evaluated relative to an equivalent configuration using AAL5 as specified in the current standards.

5.4.1 Cell Loss Ratio for CBR Transmission

This set of simulations aims at studying the influence of the packet size as well as the fraction of the link rate used by the TUT *on the CLR* observed by the receiver. Three packet sizes have been studied: 376, 752 and 1504 octets which correspond to 2, 4 and 8 MPEG-2 TS packets respectively. This is compliant with the ATM Forum specification which uses two as the default size but does not preclude bigger PDU sizes. The choice of these sizes allows to obtain the same number of cells for both AALs, respectively 8, 16 and 32, in order to be as fair as possible. Note that for the 16 and 32 cell packets, AAL5 adds 16 bytes of padding to align the data to cell boundaries.

This first set of performance experiments does not aim at studying a realistic case. It is not expected that CBR traffic will experience losses, other than those due to network outages, since the bandwidth allocation will be deterministic. Neverthelesss, the deterministic traffic pattern offered by CBR traffic allows for a relatively easy interpretation of results.

We have used the same two TUT load configurations of the precedent studies. Considering the results concerning the cell loss process, we have limited the simulations to Bernoulli background traffic sources only.



(a) TUT Load 16%   (b) TUT Load 25%

Figure 5.19: *Cell loss ratios seen by the receivers for three packet sizes for CBR traffic as a function of the background load.*

Fig. 5.19 shows the CLRs experienced by the receiver using both the MAAL and AAL5.

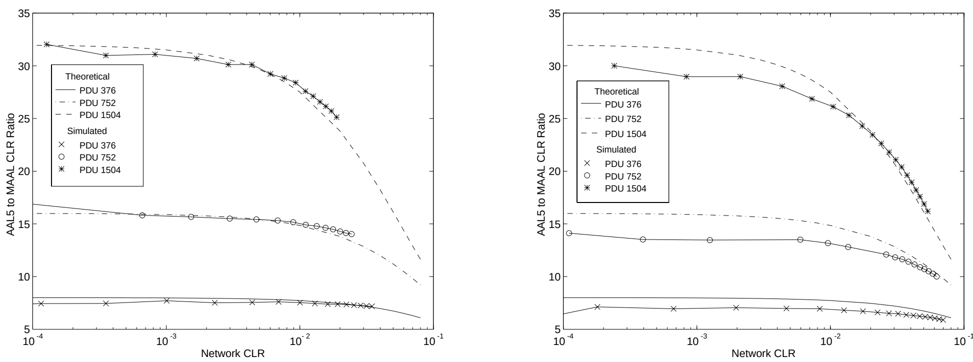
In all the experiments the CLR measured for the MAAL is equal to the network loss ratio since there is no extra discard of information at the AAL. Comparatively, the receiver using AAL5 experiences a much higher CLR due to the packet discard mechanism[1]. This behavior is explained by the uniformly distributed cell loss process. Indeed, with such a loss process the probability of observing multiple cell losses in a single packet is small therefore leading to a high number of packets discarded due to single cell losses as calculated in Sec. 4.5.4.2.

The CLR experienced by the AAL5 user heavily depends on the packet size, since the difference between both AALs increases with large packets. Both figures show that increasing the CLR slightly reduces the difference between the AALs but the multiplying factor remains nearly constant. This suggests that a clustering effect eventually appears under very high CLRs therefore reducing the influence of the packet discard mechanism.

To verify this phenomenon, we define R_{CLR} as the AAL5-to-MAAL CLR ratio actually equal to the AAL5-to-network CLR ratio. Eq. 4.1 from Chap. 4 gives the probability of losing a packet of size m cells by calculating the probability of losing at least one and at most all cells in the packet. We can therefore evaluate the number of cells lost due to the packet discard mechanism. Then R_{CLR} is given by:

$$R_{CLR} = \frac{1}{CLR} \sum_{n=1}^{m} \binom{m}{n} CLR^n \times (1 - CLR)^{m-n}. \tag{5.4}$$

Figure 5.20 depicts R_{CLR} for the six cases presented as a function of the network CLR and compares it to Eq. 5.4.



(a) TUT load 16%

(b) TUT load 25%

Figure 5.20: *AAL5-to-MAAL CLR ratio for different packet sizes.*

Fig. 5.20 shows that for small packets (8 cells), R_{CLR} remains nearly constant for all CLRs. R_{CLR} is equal to the PDU size in cells minus one. This suggests that in average, the number of cells lost per corrupted packet is close to one and behaves consistently with the iid loss process model.

When larger PDU sizes are used, the characteristic of R_{CLR} does not remain constant anymore. Two factors play a role in this behaviour: the CLR and the fraction of the link rate occupied by the TUT. When 16% of the bandwidth is used by the TUT we see that at 10^{-4}, the ratio is equal to the PDU size in cells, 16 and 32 respectively. However, the ratio as a function of the CLR constantly decreases for higher CLRs. The slope of the decay increasing proportionally to the CLR. In fact, Eq. 5.4 is a measure of the burstiness

[1]The last revision of recommendation I.363.5 allows corrupted AAL5 SDUs to be passed to the upper layers excluding incomplete SDUs (see Sec. 2.2.3.4 for details).

of the cell loss process. The smaller the value of R_{CLR} the higher, in average, the number of consecutive losses observed.

	Simulation (PDU size/load)					
	376/16	376/25	752/16	752/25	1504/16	1504/25
Max Cells Lost per PDU	4	6	7	8	7	10
CLR	3.4e-2	6.9e-2	2.44e-2	6.01e-2	1.8e-2	5.34e-2

Table 5.5: *Maximum loss parameter values.*

Figures 5.21 to 5.23 plot the distribution of the number of cells lost per corrupted packet for the three packet sizes studied. These figures capture the possibility of having *multiple non-consecutive* cell losses in the same packet. Comparing these figures to Figs. 5.7 and 5.10 shows that for small packet sizes, the behavior is relatively close. In particular, Figs. 5.22 (a) and 5.23 (a) show that for a fraction of the link rate of 16%, the number of cells lost per packet remains small. The larger packet sizes actually increment the tail of the distributions with the increasing CLRs to up to 7 cells, for both 16 and 32 cell PDU cases, in the very high CLR range (10^{-2}) as shown in Tbl. 5.5, but the impact on the overall performance is small because the percentage of *single cell losses* still predominates with values varying between 80% and 70% in the worst CLR case to close to 90% in the majority of the cases which cover a more realistic interval of CLRs between 10^{-3} and 10^{-6}.

The characteristic of Fig. 5.20 (b) is different. Even if in Sec. 5.3 we have not observed major correlations when 25% of the bandwidth is required by the TUT connection, the current results tend to prove that some exist. The experimental figures do not match the calculated ones for R_{CLR}. This indicates that some clustering occurs in the cell loss process. Figs. 5.21, 5.22 and 5.23 (b) show that the proportion of consecutive losses has increased compared to the 16% load case. This explains the difference between the theoretical and simulated results as well as the important decay observed in Fig. 5.20 (b) which confirms the *major influence* that the packet size has onto the CLR experienced by an AAL5 user.

In summary, the results obtained show, given the results of Sec. 5.3, that for a CBR cell flow representing a relatively small fraction of the link rate, the probability of having more than one cell lost per packet is small. This is particularly true for small packet sizes (up to 16 cells in our tests). Consequently, an AAL using packet discard performs poorly because it significantly increases the CLR experienced by the user. This implies also that *the packet size*, which does not depend on the AAL but on higher layers, *has a major impact on the AAL performance*. The bigger the packet the worse the performance. On the other hand, the granularity of the MAAL makes this layer *insensitive to the packet size* as well as to the *burstiness of the cell loss process*, explaining the performance gains achieved. The comparative simulations show that the MAAL performs much better than AAL5 achieving smaller CLRs for all the situations studied.

5.4.2 Cell Loss Ratio for VBR Transmission

Given the precedent results, we have limited the simulations for VBR background traffic to a single PDU size of 376 octets. The same network configuration has been used as well as the same three VBR video streams described in Sec. 5.3.2.

The curves shown in Fig. 5.24 illustrate again the improvement achieved by the segmentation mechanism of MAAL. Regardless of the burstiness of the foreground traffic, AAL5

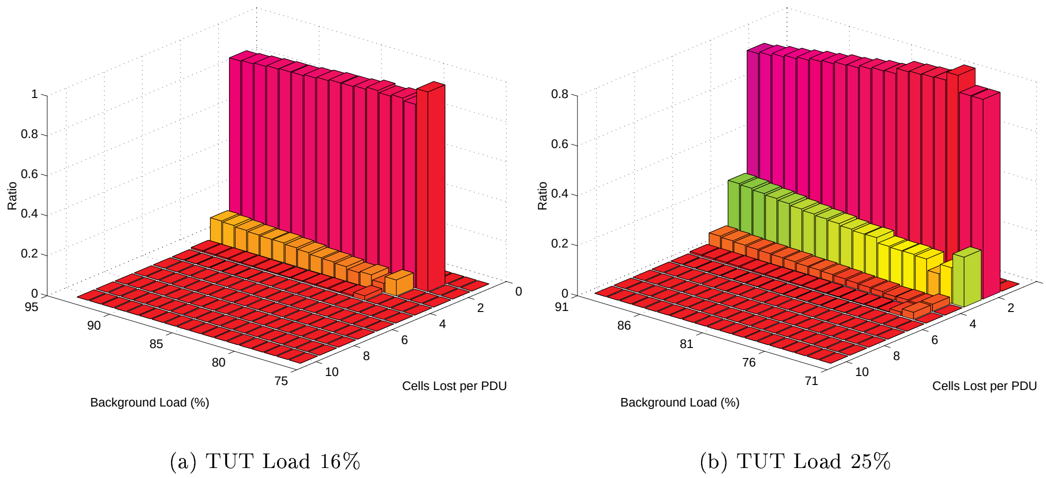


Figure 5.21: *Distribution of cells lost per packet for a PDU size of 376 octets.*

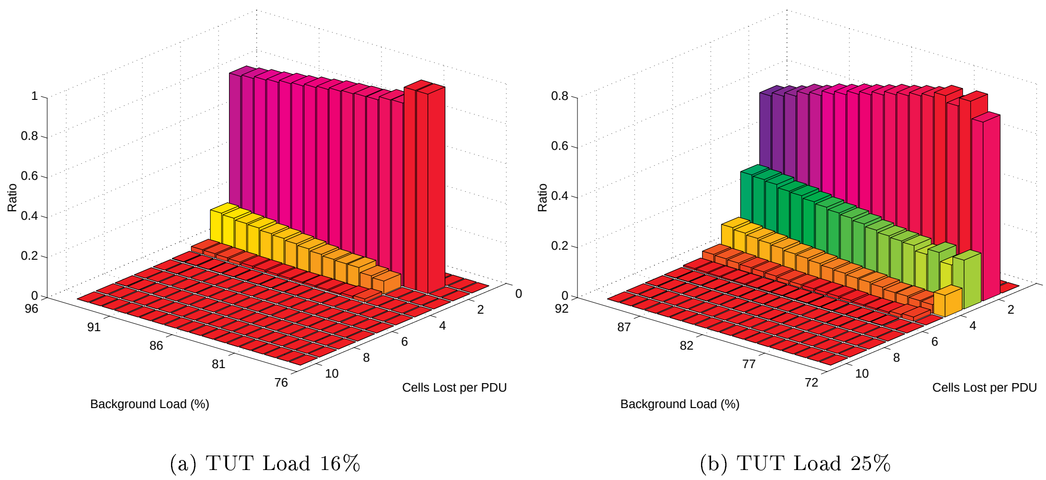
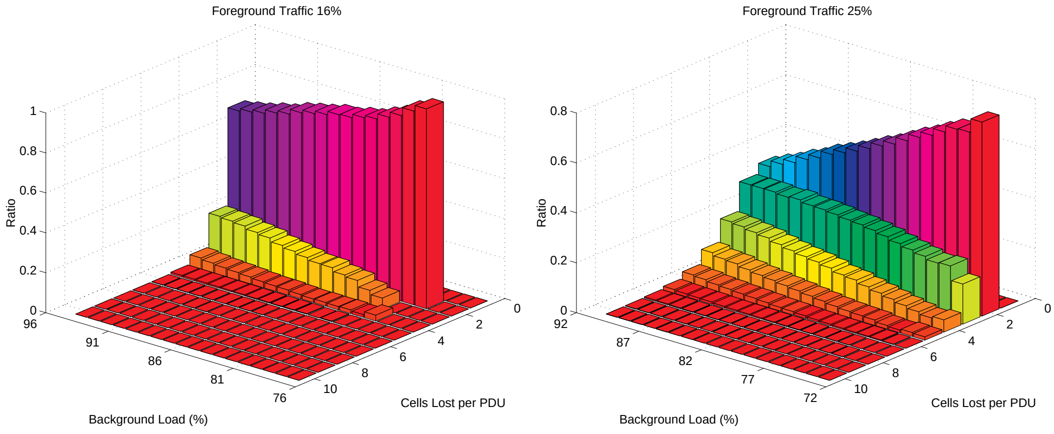


Figure 5.22: *Distribution of cells lost per packet for a PDU size of 752 octets.*

always gives worse CLR figures than the MAAL. Since the cell loss process has been shown to be almost not correlated for these configurations, the results look similar than for the CBR case previously presented.

Again, as depicted in Fig. 5.25 the distribution of the number of cells lost per packet explain the weak performance of AAL5 for the transmission of video. Even if the increased burstiness of the foreground traffic entails an increase on the cell loss burst sizes and on the number of cells lost per packet, in average, the MAAL still achieves better performance than AAL5. Note that for the highest quality video which at peak rate uses more than 30% of the link bandwidth, the maximum number of cells lost per packet summarized in Tbl. 5.6 is equal to the size of the PDU which means that full packets are lost. This should reduce the impact that the packet discard mechanism has in comparison to the MAAL mechanism. However, the percentage of complete lost packets is very small and hence has almost no impact on the comparative CLR figures.



(a) TUT Load 16%

(b) TUT Load 25%

Figure 5.23: *Distribution of Cells Lost per Packet for a PDU size of 1504 Octets*

These results, however, must be used with caution. The main difference between the CBR and VBR traffic profiles resides in the validity of the first order statistics. Since CBR traffic always has a periodic and therefore predictive behavior, mean values are generally good estimators of the measured event. This is not always the case for VBR traffic. Average values show a trend of the estimators used, but they fail to catch instantaneous behavior. As an example, consider the burstiness factors calculated in Tbl. 5.3. Even if the stream compressed with the quantization factor of 52 gives the higher burstiness figure, the highest quality video has the highest variations in traffic (the peak-to-minimum ratio).

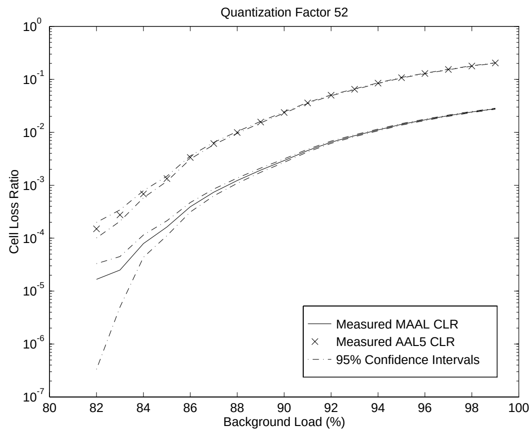
In other words, even if the obtained figures show that in average the MAAL performs better it is possible that in high cell rate periods both AALs achieve similar performance while in low cell rate periods the MAAL will behave better. Whether the impact of cell losses is more annoying in high or low source activity periods is difficult to say and requires other metrics to be discussed in Sec. 5.4.4.

Q Factor	14	28	52
Max Cells Lost per PDU	8	6	4
CLR	1.35e-1	3.75e-2	1.41e-2

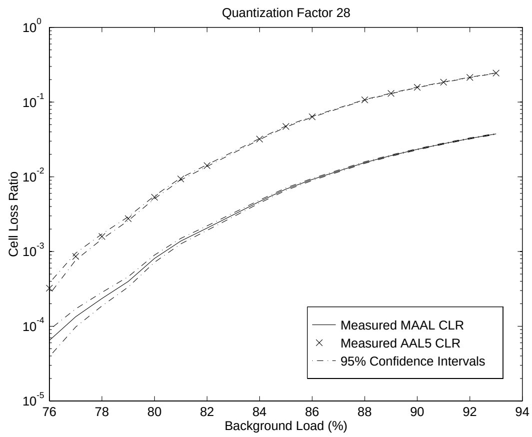
Table 5.6: *Maximum number of cells lost per packet for VBR video streams.*

In summary, the utilization of the proposed multimedia AAL gives better figures in terms of network QoS experienced by the receiver than an equivalent transmission over AAL5. We have observed that for different types of foreground and background traffic, a packet-oriented AAL such as AAL5 including the packet discard mechanism leads to CLRs almost an order of magnitude highers compared to a cell-oriented AAL. The benefit of using sequence numbers and passing corrupted packets to the upper layers is that CLRs several times lower could be achieved by introducing a small overhead per cell. In addition, the cell-oriented mechanism is not sensitive to the size of the packets used for transmission while a packet-oriented mechanism with packet discard is.

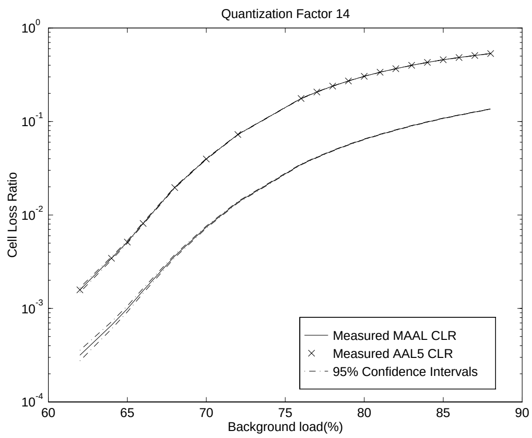
A slight performance degradation compared to the CBR case appears as expected due to a short term cell scale burstiness which is not captured by the estimators used. Albeit the



(a) Quantization Factor 52

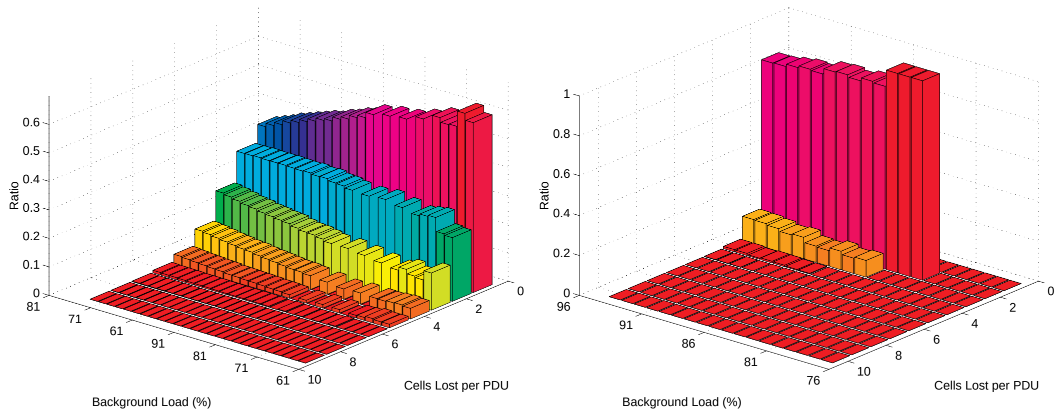


(b) Quantization Factor 28



(c) Quantization Factor 14

Figure 5.24: *Cell loss ratios experienced by the receivers for VBR traffic as a function of the background load.*



(a) Quantization Factor 14

(b) Quantization Factor 52

Figure 5.25: *Distribution of cells lost per packet for a PDU size of 376 octets.*

results of Sec. 5.3.2 show that this phenomenon is negligible, since no correlations appear in the cell loss processes, it is not possible with network metrics to evaluate the impact that such losses have onto video. The introduction of perceptual quality metrics in the next chapter will show how this cell scale burstiness affects the video information.

Given the results obtained, a sequence number field of 5 bits is enough to cover the majority of the cell losses. The results show that actually 4 bits should be enough since we have not observed any burst of 16 lost cells. However, to be robust enough 5 bits are preferred. Also, the function allowing to pass corrupted packets, with dummy cells inserted or not, heavily improves the CLRs observed by the AAL user compared to AAL5's performance.

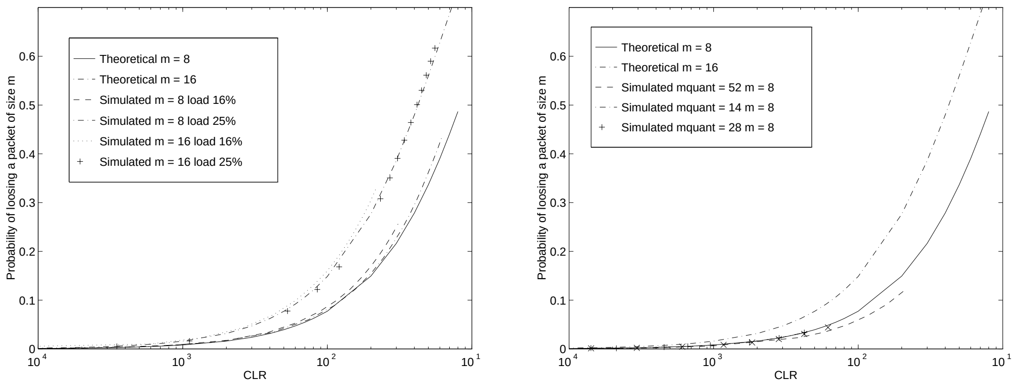
To conclude, this experiments show that a stream-oriented AAL performs much better than the currently used AAL5 under the constrains posed by real-time interactive multimedia applications.

5.4.3 Cell Loss Recovery Performance

Since the low traffic source assumption has proven to be a good approximation, the calculations made in Sec. 4.5.4 let foresee an important improvement in CLR values if the cell level FEC is used. We present in this section a comparative performance study of both cell and packet level FEC mechanisms covering both CBR and VBR traffic.

Figure 5.26 compares the packet loss probabilities calculated in Chap. 4 using Eq. 4.1 to the simulation results obtained with the same configurations described in the precedent sections for both CBR and VBR traffic.

As expected, given the results of the precedent sections showing the accuracy of the cell loss iid model, both figures match very well the theoretical curves. Only under very high CLRs, some deviations appear, in particular for VBR traffic, mainly due to the clustering effect in the cell loss process. However, the low correlation figures observed entail that the packet loss ratio curves slightly deviate from the random case. This figures bring to the fore the increased loss a user experiences when a packet discard mechanism is used.



(a) CBR Foreground Traffic

(b) VBR Foreground Traffic

Figure 5.26: *Packet loss ratio vs. cell loss ratio for VBR TUT.*

Since AAL5 does not perform FEC, we have implemented a packet based FEC on top of AAL5. As shown in Fig. 5.27, k data packets are protected by h redundancy packets, which lead to a FEC block of size $k+h$. The FEC scheme is based on the RSE codes described in Sec. 4.5.4.2.

The protection scheme has been set to achieve the same code rate for both cell and packet level FEC mechanisms. The AAL5 transmission protects k packets with a single FEC packet ($h = 1$) of size m cells. The MAAL scheme generates m FEC cells that protect all $m \times k$ cells of data. Setting $k = 8$, the overhead *per FEC block* is:

$$\begin{aligned} overhead &= \frac{1}{k+1} \\ &= 11\%. \end{aligned}$$

To evaluate the efficiency of both mechanisms, we define the random variable X as the number of packets received in a FEC block. Therefore, the recovery efficiency is given by the probability of receiving at least k among $k+h$ packets expressed as $P(X \geq k)$.

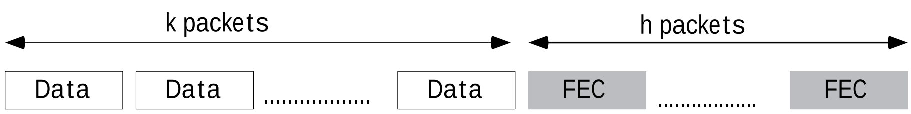
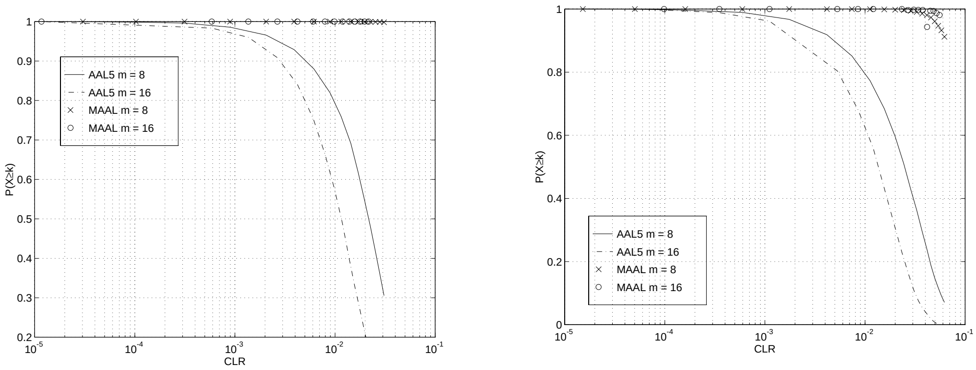


Figure 5.27: *Packet-based FEC correction mechanism implementation.*

As shown in Fig. 5.28, which plots $P(X \geq k)$ for CBR traffic, the cell level protection mechanism is much more efficient for all packet sizes and all TUT loads for the same overhead.

Also, as expected, Fig. 5.29 shows that when reducing the overhead of the cell protection mechanism by a factor m, the recovery efficiency is equivalent to the one achieved with AAL5 using a full FEC protection packet.

Fig. 5.30 depicts the recovery efficiency of the same cell and packet based protection schemes for the three different quality bitstreams. As the precedent figure suggests, even



(a) TUT Load 16%

(b) TUT Load 25%

Figure 5.28: *Comparative loss recovery performance, $P(X \geq k)$, for transmission over AAL5 and MAAL with two packet sizes. CBR foreground traffic.*

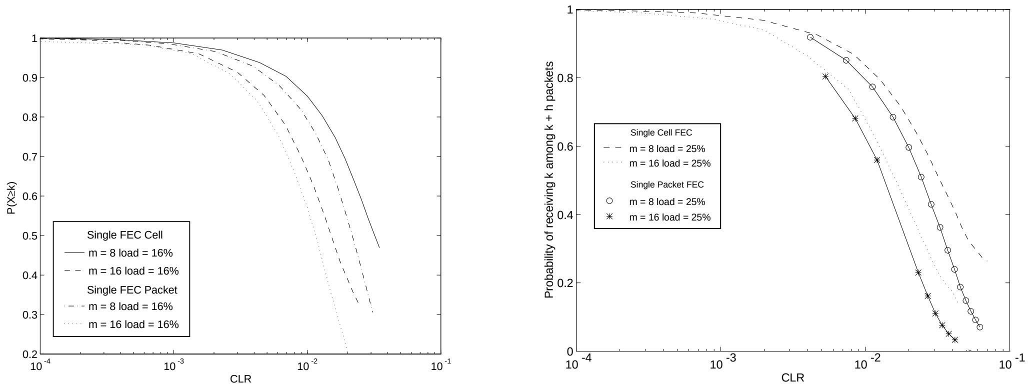
for VBR traffic the cell based FEC mechanism is much more efficient than the packet based one. However, it is clear that when losses occur while transmitting at peak rate the FEC mechanisms loses part of its efficiency. The simulations of Sec. 5.4.2 showed that the probabilities of having large bursts of cells lost is quite low. Fig. 5.25 shows that even if up to 12 consecutive cells could be lost, for the highest quality stream, and that full PDUs could be lost, the probabilities of these occurrences are very low and for extremely high CLRs. The fact that the cell based mechanism is able to recover up to eight cells in a FEC block regardless of their *burstiness* explains the better performance achieved by the MAAL.

In summary, the cell level recovery mechanism performs much better for an equivalent overhead for both CBR and VBR types of traffic. In addition we show that reducing the overhead by a factor m, m being the size of the packet in cells, we achieve the same recovery performance than a packet based error correction mechanism. The reasons to this improvement are twofold: first, the cell loss process is almost not correlated which leads to a worst case scenario for the packet based mechanism. Second, the cell level mechanism is neither sensitive to the burstiness of the cell loss process nor to the packet size.

However, the overheads due to FEC are not negligible and do not take into account the impact that the loss of the data has onto video. Since we have proposed the FEC mechanism to be selective, we propose in the next chapter a solution that will make use of this feature to reduce the overhead and will also take into account the nature of the data.

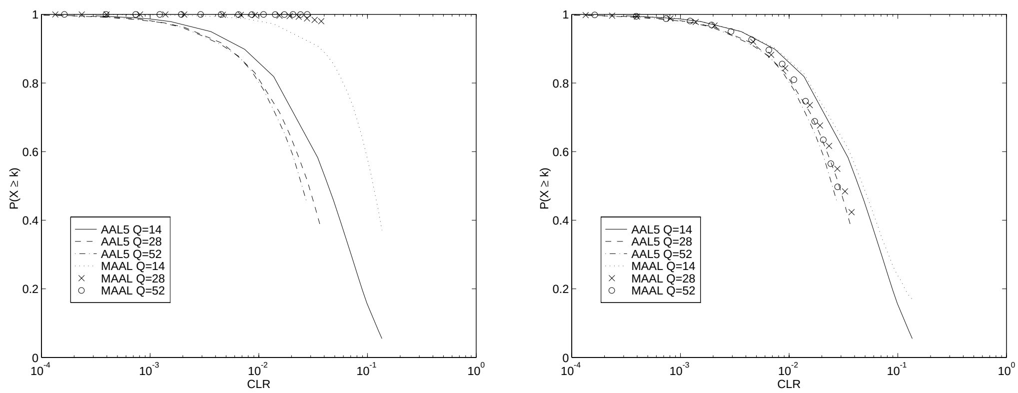
5.4.4 Perceptual Quality Performance

We have proved via simulation that a stream-oriented AAL performs several times better from the network perspective than a packet-oriented AAL such as AAL5. However, there is no direct mapping between this performance improvement and the quality perceived by the user. Due to the different types of data carried in a multimedia stream, the impact of loss may be different and in general unpredictable. This is particularly true for compression algorithms using motion estimation and other prediction techniques such as MPEG-1, MPEG-2 and H.261. To overcome this problem, we have evaluated the video quality obtained from a user perspective using a *perceptual quality metric* described in Sec. 3.2.2.



(a) Foreground Load 16% (b) Foreground Load 25%

Figure 5.29: *Comparative loss recovery performance, $P(X \geq k)$, for transmission over AAL5 and MAAL with different FEC overheads. CBR foreground traffic.*



(a) same overhead (b) different overhead

Figure 5.30: *Comparative loss recovery performance, $P(X \geq k)$, for transmission over AAL5 and MAAL with three quantization factors. VBR foreground traffic.*

The results presented here cover a subset of the simulations described earlier for both CBR and VBR foreground traffic. We, in addition, do not present in this section any confidence interval. The reason is that the evaluation of a single point in the curve takes approximately three days of processing time for a 1000 frames sequence in a 160MHz bi-processor SUN Sparc 20 workstation with an optimized C code. More general results with confidence intervals,including the performance of the selective FEC mechanism, will be given in the next chapter .

Figure 5.31 shows the NVFM quality assessment as a function of the CLR for the transmission of CBR MPEG-2 encoded streams over AAL5 and the MAAL. Both curves show two regions: a first one with very small variation in quality and a second one where the quality drops very quickly. In the first region, both curves are nearly horizontal and very close to the reference value. There is a very small variation in the quality due principally

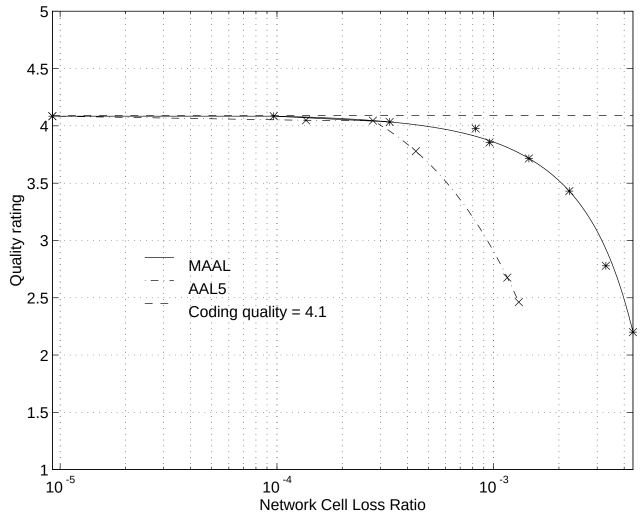


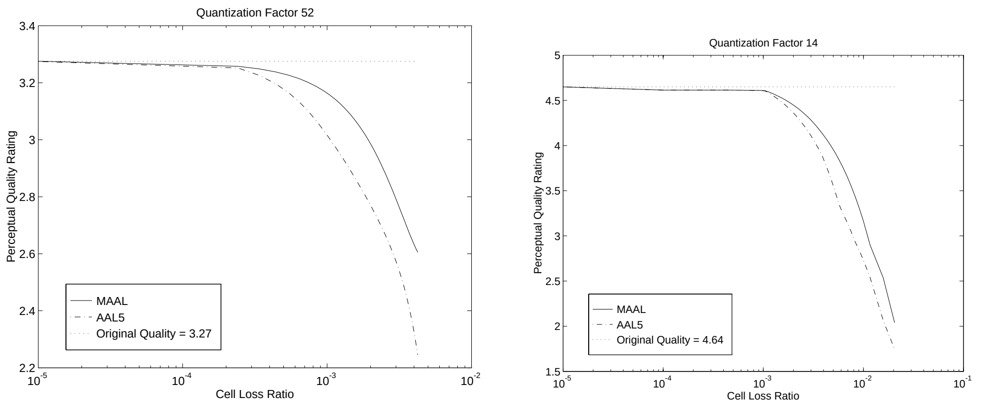
Figure 5.31: *Quality rating vs. network load for AAL5 and MAAL for CBR transmission. PDU size 376 octets TUT load 16%.*

to relatively sparse cell losses that have low impact in the image. The second region where both curves have a steep slope show that the increased number of cell losses begin to have a heavy impact onto the image quality. This decay is basically due to the loss of large amounts of data that include syntactic information such as, in the case of MPEG-2, PES or frames. It is interesting to note that the critical CLR is not the same for AAL5 and the proposed AAL. AAL5's critical point appears for a CLR of 3×10^{-4}, while the MAAL's critical value appears at 10^{-3}. In fact, for a given loss ratio beyond the critical point, we achieve a significant gain (close to 1) in terms of perceptual quality. This has been obtained by *applying only the proposed segmentation mechanism and dummy cell insertion* which clearly shows the influence the AAL5 packet discard mechanism has onto the perceived quality.

The perceptual quality achieved for the transmission of VBR video streams is depicted in Fig. 5.32. Unlike the cell loss figures showed, bursty traffic has a clear influence on the perceived quality. The performance achieved by the MAAL proves slightly better in the region $10^{-4} - 10^{-3}$ however, for higher CLR values both figures converge again. This behavior is explained by the better granularity of the MAAL which allows to pass incomplete data packets to the upper layers. The lack of such possibility in AAL5 penalizes its performance from a perceptual point of view as shown in Fig. 5.33. The perceptual quality of the AAL5 video stream drops to lower values than the MAAL video when errors occur leading to a lower average value. This is due to the fact that the regions affected by errors in the AAL5 stream are bigger. As shown in both Figs. 5.33 (a) and (b) the propagation of such larger errors leads to more important quality drops in consecutive frames.

The reason to the converge of the curves under very high cell losses is double. Firstly, with increased network load, the cell loss process becomes more bursty and groups the losses into small clusters reducing the negative effect of AAL5's packet discard mechanism. Secondly when large cell loss ratios occur, the pictures are so severely damaged (including lost frames) that the perceived quality drops very fast for both AALs. However, the MAAL proves to be more reliable without any extra overhead in a large range of cell loss ratios.

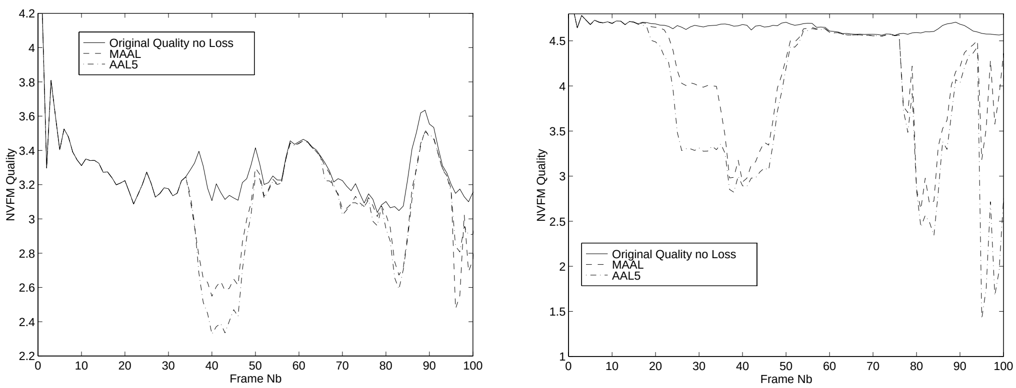
A very interesting result shown by these figures is that unexpectedly, the perceptual quality remains close to the original value under relatively high CLRs generally considered as unacceptable for any type of service. Therefore, the application of *statistical multiplexing* appears as an interesting way of using this phenomenon. The utilization of the MAAL as shown in the figures may lead to even better network utilizations due to its better robustness to network errors.



(a) Quantization factor 52

(b) Quantization factor 14

Figure 5.32: *Quality rating vs. network load for VBR transmission over AAL5 and MAAL.*



(a) Quantization Scale Factor 52

(b) Quantization Scale Factor 14

Figure 5.33: *Per Frame Quality Rating vs. Time for VBR Transmission over AAL5 and MAAL for two Encoding Qualities*

5.5 Multimedia AAL Performance: Summary of Results

We have shown throughout this chapter that the low traffic source assumption which leads to a uniformly distributed loss occurrences process accurately approximates the cell loss process observed in the simulations.

The scenarii studied assumed a large number of sources generating the background traffic. In addition, due to the delay constraints of interactive multimedia applications we have chosen small buffer sizes for the switching elements. Finally, the majority of the observed TUT data flows used in average less than 10% of the link bandwidth which is in general the case for MPEG-2 encoded video.

The MAAL-SAR performance studies developed in this chapter show that from a networking point of view, the utilization of a cell-oriented segmentation mechanism such as the one proposed for the MAAL gives better figures in terms of QoS experienced by the receiver. We have observed that for different types of foreground traffic, a packet-oriented AAL such as AAL5 including the packet discard mechanism leads to CLRs almost an order of magnitude highers compared to a cell-oriented AAL. The benefit of using sequence numbers and passing corrupted packets to the upper layers is that CLRs several times lower could be achieved by introducing a small overhead per cell. Moreover, the cell-oriented mechanism is neither sensitive to the size of the packets used for transmission nor to the burstiness of the cell loss process while a packet-oriented mechanism with packet discard is. In fact, if correlations and therefore consecutive losses appear, then the performances of the MAAL will drop but never below the AAL5 performance.

We also show that a cell based protection scheme based on RSE FEC is much more effective than the equivalent mechanism applied on top of AAL5. The MAAL achieves equivalent recovery performance with a fraction of the overhead, roughly $\frac{1}{m}$ m being the size of a packet in cells. This is explained by both the characteristics of the cell loss process which is the worst case for a packet-oriented AAL and the good performance of the MAAL regardless of the packet size and the burstiness of the cell loss process.

Chapter 6

Network Adaptation Layer for Interactive Multimedia Applications

6.1 Introduction

This chapter presents the design, description and testing of the services and functions proposed for a *new multimedia-oriented Network Adaptation Layer* (NAL). The NAL task is to adapt as good as possible the data to be transmitted to the network.

The chapter begins with a rationale for a NAL. The NAL's purpose is to provide means to perform intelligent data protection. This may be performed only if NAL's are codec specific. The advantages and drawbacks of this specific versus a more generic approach for the NAL design are then discussed. We then derive a set of design principles we consider mandatory to build NALs. On this basis, we develop the generic functions to be provided by the NALs. This functions are then applied to develop an MPEG-2 specific NAL, including a perceptual data protection mechanism. We study the perceptual impact that such a protection mechanism has. We finally present the benefits of using such layer in terms of network and perceptual QoS.

6.2 Rationale for a Network Adaptation Layer

An AAL by definition has to provide generic functions for a given class of service. But to fully exploit the functionalities of the AAL, it is necessary to have a layer that will efficiently interface the network and the applications. This role is devoted to the Network Adaptation Layer.

The concept of network adaptation is not totally new. ITU-T recommendation H.222.1 [128] describes a network adaptation layer *specific* to H.262 [19] and H.263 [21] video codecs (see Sec. 3.5.1). This specification does not clearly define a layer but a set of functions and services for the multiplex and synchronization of audiovisual communications. It specifies the peer-to-peer syntax, semantics and interactions with AALs 1 and 5 which automatically places these functions on top of the AAL. Among the services provided by H.222.1 we find:

- multiplexing of data streams carrying only one media source type
- timebase recovery by providing additional timestamping

- error reporting to the H.222.1 user
- timing jitter removal based on H.222.1 time stamps.

In [117, 118] a similar concept is discussed. The authors propose a *channel adaptation* specific to MPEG-2. This adaptation is not exactly defined as a layer but its purpose is to do an *intelligent* mapping of the upper layer SDUs to AAL-SDUs. However, the level of integration proposed requires to totally modify the AAL to integrate the functions making the AAL MPEG-2 specific which is against the definition and functions of an AAL.

The results of Chap. 5 have shown that significant improvements in terms of user and network quality of service can be achieved by using FEC-based data protection. However, such improvements require additional bandwidth due to the overhead necessary to accommodate the extra FEC information. In addition, given the structure of audiovisual information and the loss propagation phenomena characteristic of compression (see Secs. 2.1.2.1 and 3.2.2), it is obvious that a systematic or *blind* data protection is not optimal in terms of user perception.

We propose in this work to push further the codec-specific network adaptation concept to optimize the interface between the application and the underlying network and to achieve a reliable but low overhead transmission of video streams by means of a *selective protection mechanism*. Since this requires an *a priori* knowledge of the information structure to be transmitted, the proposed NAL will be *single codec* specific.

6.3 Generic versus Specific Network Adaptation

As we have described in Sec. 4.3, the most used audiovisual compression algorithms to date have some common characteristics, such as the organization of the data in hierarchical structures prefixed with headers, that allow to define a set of generic functions for the transmission of multimedia data over ATM networks. However, the approach taken to organize the compressed data into syntactic and semantic components imply that every codec has its own unambiguous set of codewords that define the syntax. Consequently, it is not possible to achieve a generic protection scheme which will not be systematic. If, as suggested, FEC techniques are used, then the overhead generated by the recovery scheme is far from being negligible if an efficient protection scheme is targeted. In addition, since the importance of the various types of information is very different it becomes hard to justify a uniform protection scheme.

If, instead, a specific rather than a generic approach is considered then protection schemes tailored to given applications or sets of applications could be developed, achieving much more reliable data streams because targeting sensitive data. Therefore, issues such as the uniform protection scheme could be solved while simultaneously reducing the overhead.

From a protocol point of view this approach is howsoever not the rule of thumb. The 7-layer protocol stack of the OSI model has clearly been defined to avoid such specific layer issue that otherwise would lead to a large number of layers each one designed to fit a type of applications. Reality has shown that beyond the transport layer almost no further requirements have been identified to fill the session and presentation layers.

To solve the incompatibility between the rigid structure of the OSI protocol model and the diverse needs of applications, Application Level Framing (ALF) was proposed as an alternative protocol model [158] (see Sec. 3.5.4).

The approach taken in this work is between both trends. While ALF meets the criteria of knowing the syntax and semantics of the data to be transmitted, because integrated into the application, it has the disadvantage of not adapting the data to the underlying network. Conversely, if the OSI model is used it is not possible to provide a semantics-based protection mechanism.

The Network Adaptation Layer proposed in this chapter performs application or codec-specific protocol tasks while remaining generic enough to be designed as an independent layer. The reasons behind our proposal are the following:

- multimedia applications share a set of common elements for coding audio and video information and can therefore be grouped into a relatively small number of categories (e.g. by codec). Thus, the NAL cannot be considered as an integral part of an application because a certain genericity at this level exists. Still, the NAL cannot be considered as a classic protocol layer because it depends on codec functions and therefore cannot be designed as a standard protocol layer that should work regardless of the application

- traditional protocol functions such as the multiplexing of different applications data flows which may become increasingly important in multimedia communications obviously cannot be handled by one application. It is not a function performed by the transport or network layers and therefore has to be performed above these

- the development of the Java language allows for the development of dynamically downloadable pieces of executable code. This paves the way for the development of dynamic high level protocol layers or *'proclets'* downloadable into the receiver's terminal. Proclets would add high level protocol functions required by specific applications to improve network services.

The NAL targets interactive real-time multimedia services. These services share a set of common requirements independent of the encoding scheme used. This calls for the definition of a set of design principles that should be shared by any NAL even if the implementation of the functions themselves is codec specific. These design principles described in the next section are:

- low delay: due to the nature of the data and the requirements of both the application and the user

- low overhead: the NAL should provide an intelligent data protection scheme that simultaneously minimizes the overhead and achieves a high degree of reliability in terms of loss impact onto the perceived quality

- modularity: the NALs have to be codec dependent and self-contained. This means that if an application uses a different codec, it has to be possible to use the correspondent NAL without any other modification of the protocol stack or the application.

6.4 The Three Design Principles

6.4.1 Low Delay

To achieve an efficient transmission of interactive multimedia data streams, it is necessary to generate as low delay as possible. The NAL functions must be simple in order to keep the processing delay to a minimum. In addition to the processing delay, the accumulation of data packets to create the NAL-SDUs may in some cases add extra delay jitter.

The results of Sec. 5.4.3 show that the MAAL recovery FEC efficiency does not directly depend on the size of the packets. The recovery level depends on the code rate $\frac{h}{k}$ where h is the number of data cells and k is the number of FEC cells. If this code rate is fixed then the smaller the number h of data cells, the smaller the number k of FEC cells required. It is therefore a good guideline to generate small size packets.

Last but not least, unlike AAL5, the MAAL overhead per packet is constant. The convergence sublayer does not add any extra information. The SAR uses an octet from the cell payload for control information but this overhead does not depend on the size of the PDU to transmit. Therefore, using small packets does not affect the next design principle which targets a low overhead.

The drawback of such guideline is that for codecs such as JPEG which encode entire frames as a single entity it would be necessary to segment the packets which will therefore add extra overhead.

6.4.2 Low Overhead

The results obtained in Chap. 5 show that the cell loss processes observed for the transmission of both CBR and VBR video streams allows for an efficient utilization of FEC recovery techniques.

However, the protection of data by open-loop techniques such as FEC always comes with the drawback of adding extra overhead. The nature itself of the technique does not allow to have a null overhead. This effect can howsoever be reduced to small values if the nature of the data to be transmitted is taken into account and also if we consider that audiovisual communications accept limited loss. Considering the separation of audiovisual information into syntactic and semantic components, a low overhead, albeit, reliable multimedia flow can be achieved by selectively protecting the most relevant data from a perceptual point of view. By most relevant data we mean the data whose loss has the heaviest impact onto the displayed material. In general the amount of syntactic information is small compared to the raw video information. Therefore selectively protecting this data generates small overhead while still being efficient from a perceptual point of view.

6.4.3 Modularity

One of the key features of NALs is their possibility to identify the type of information contained into the packets. Since this depends on the codec who performs the compression, the NALs have to be codec dependent. Therefore it is necessary for these layers to be modular and self-contained. This means that for a given application and protocol stack it has to be possible to change the NAL without any other modification of the application or the protocol stack

Different NALs are envisaged to be implemented on top of the MAAL to handle the different encoding algorithms as depicted in Fig. 6.1. Moreover, the general adoption of Java could lead to the development of dynamic protocol layers. The concept of *applet* applied to protocol layers, the *proclets*, would allow to load in a per-connection basis the additional functionalities required by a certain type of codec. Since proclets would be considered as standard applications by the terminals they would have to run in user space. To achieve high speed and enough performance to handle high bit rates it would be mandatory to keep NAL functions as simple as possible.

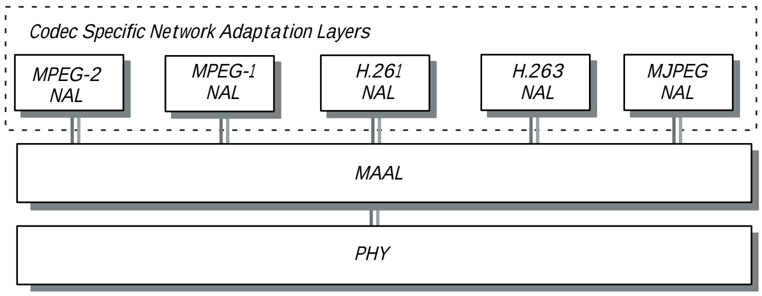


Figure 6.1: *Two layer reference scheme.*

6.5 Proposed Common Functions for a Network Adaptation Layer

The NAL provides a set of services to the layer above, the application. To support the services, the NAL has to provide a set of functions that must be *common* to all NALs. Even if the NALs are codec specific, the different layers must provide the same functions. The differences will reside in the way, these functions are performed which will be tailored to a given codec syntax.

The functions we propose for the NALs to provide are:

1. segmentation and reassembly of user information

2. selective data protection.

Other functions such as multiplexing and synchronization, as described in H.222.1, are not precluded. However, this thesis does not cover such functions.

6.5.1 Segmentation and Reassembly

The MAAL does not perform any padding or delineation because it provides the transfer of fixed length packets. The overhead is constant. Therefore it is the NALs task to provide such delineation.

The current compression algorithms (see Sec. 2.1.2.1) deliver both fixed (e.g. MPEG-2) and variable (e.g. JPEG) size packets or Applications Data Units (ADUs). Therefore, the NAL has either to perform a segmentation of ADUs into smaller fixed-size data units that the MAAL could deal with or conversely to assemble them into bigger NAL-SDUs. These functions require from the NAL to provide a specific packet structure containing delineation fields and a padding function. These implies that the NAL could add some overhead.

The rationale for leaving the delineation function to the NAL is that the overhead that may be generated at this layer could be minimized due to the knowledge of the data structure. The NAL could use the data headers as the delineation flags. Still some overhead should be added when data alignment to cell boundaries becomes necessary. Also, if small packets are used this overhead should be small.

In summary, since the NAL is specific to a given codec it could take advantage of the structure of the information to perform such tasks as packet segmentation and reassembly, packet delineation and alignment to the required boundaries, therefore introducing very low or even no overhead at all.

6.5.2 Selective Data Protection

The flexibility of FEC data generation offered by the MAAL allows to have different protection schemes without any modification of the functions of the AAL. The NAL could provide different protection algorithms depending on the degree of reliability required by the application or on the channel to be used.

We describe in the next paragraphs different protection mechanisms that could be implemented at the NAL.

6.5.2.1 Perceptual Syntactic Information Protection (PSIP)

As it has already been described in precedent chapters, the impact of data loss depends on several factors, the most common being the loss of syntactic information. Since the loss of a header entails the impossibility to decode the encapsulated information, even if it is available, headers are significantly more important than raw video data.

Another important factor that influences the perceptual impact of data corruption is the type of picture hit by the loss. This is valid for encoders using predictive algorithms such as MPEG1 and 2 and H.261. Due to the organization of pictures into GOPs and to the temporal prediction, some losses are subject to temporal propagation and therefore have a large visual impact.

Based on these perceptual properties of compressed audiovisual information it is possible to derive a protection mechanism. We call this technique *Perceptual Syntactic Information Protection* (PSIP). The principle of this technique is to selectively protect the data headers of a multimedia stream.

The advantage of this protection mechanism is that it is easy to apply to any kind of hierarchically structured audiovisual information. It is very simple to implement and does not require many processing time. Moreover, it is extremely flexible and could even be dynamically changed during the transmission if network conditions change (given that a feedback channel exists). However, this is not a major need and could introduce unnecessary complexity in multiparty configurations.

6.5.2.2 Perceptual Quality Information Protection (PQIP)

A PSIP protection scheme actually protects perceptually relevant information because the loss of the protected headers has the heaviest impact onto the displayed material. However, a PSIP protection does not take into account the *perceptual relevance* that a part of the image has to the user. If the application is able to indicate to the NAL the perceptual relevance that the information to be transmitted has it is possible to apply the required protection on this basis. In [166] a Perceptual Visibility Predictor (PVP) able to calculate the perceptual relevance or perceptual activity of the different parts of an image based on the MPQM perceptual quality metric is developed. Based on this perceptual relevance it is possible to develop an algorithm which will prioritize the information on this basis. We call this technique *Perceptual Quality Information Protection* (PQIP).

The drawback this technique has is that it involves more calculations than the PSIP. Also, the calculations have to be done in the original image, therefore, the NAL is not able to perform them since the NAL handles compressed information and not the original data. It would be necessary to modify the codec in order to pass the perceptual relevance of the data to the NAL. Then the NAL would ask for the required protection according to the priority scale used. It is clearly more difficult to implement. The problem that PQIP has is that by working at the image level it completely bypasses the relevance of the syntactic information and gives only information about the raw video relevance.

6.5.2.3 Source Rate Based Information Protection (SBIP)

Chapter 5 has experimentally shown that for VBR communications a small correlation exists between the CLR and the source bit rate. Due to the variations in bitrate of the video source, the network utilization increases during periods of high cell rates. Therefore the probability of cell loss is increased during these periods of high activity. Hence, a protection technique based on a monitoring of the source bit rate could help to reduce this cell loss probability during the specified high cell rate periods. This technique we call *Source Rate Based Information Protection* (SBIP) should be used as an enhancement to both PSIP and PQIP protection schemes. Indeed, the SBIP scheme could even reduce the overhead by intelligently reducing the number of FEC cells to be generated during low bit rate periods.

6.5.2.4 Summary

The concept of Network Adaptation Layer presents many advantages. It is very flexible because different protection algorithms could be used. In addition, if Java applet implementations of these layers are done in the form of proclets, there is no need to know if the receiver has or not the protocol layer. The NAL could be transmitted and dynamically integrated in the stack with the supplementary advantage that the end-user does not need to have a large set of NALs implemented to take advantage of the features offered by the NAL protection scheme. This, in addition, has the advantage of not requiring any modification of the existing decoders or set top boxes.

A NAL being codec dependent can easily take advantage of the structure of the data to be transferred to perform such tasks as packet segmentation and reassembly, delineation and boundary alignment.

We have described two perceptually-based protection schemes, the PSIP and the PQIP. The first relies on a simple concept. The multimedia data being structured and organized into syntactic and semantic information it is simple to identify this syntactic information and provide for protection of this perceptually relevant data. The second technique is more complex because it relies directly on the user perceived relevance of the image. In addition, it requires some codec modifications. One of the questions about this technique is if it efficiently protects from cell losses. It is clear that the syntactic structure of the data is not taken into account by the PQIP and therefore has to be considered as a complementary technique rather than an alternative.

Finally, the SBIP technique aims at protecting the data based on the source rate. It is an enhancement to PSIP and PQIP techniques in the sense that it fails to capture any specific data structure. However it may reduce the overhead during low bit rate periods.

Among these three techniques the only one to fulfill the modularity design principle of NALs is the PSIP. The other two techniques require either supplementary codec information or a modification of the codec itself.

The utilization of a NAL layer does have the advantage of being independent of its peer entity. That is, the receiver's NAL does not need to know which protection algorithm is being used since the protection and recovery of the data is actually done at the cell level.

Last but not least, as shown in Fig. 6.2, the NAL seamlessly integrates into the current protocol stack framework for the transmission of MPEG-2 real-time audiovisual multimedia applications over ATM networks.

The next section describes our proposal for an MPEG-2 specific NAL that will provide segmentation and reassembly mechanisms as well as selective data protection.

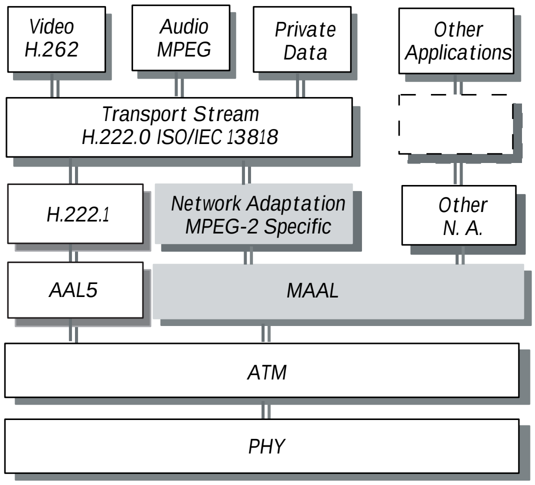


Figure 6.2: *Current ATM protocol stack and proposed layers for the transmission of real-time multimedia communications.*

6.6 MPEG-2 Network Adaptation Layer Proposal

The data generated by an MPEG-2 application could be delivered in two formats; Transport Streams and Program Streams. The former are packets of fixed length of 188 bytes and are aimed at transmission over error-prone environments such as packet networks. The latter are packets of variable length and are aimed at Digital Storage Media (DSM) or transmission over error-free environments. The current specifications for transmission of MPEG-2 video over ATM networks only consider the TS packet format for transmission [161] [167]. Recently, the development of the Digital Versatile Disk technology (DVD) has raised in the ATM Forum Audiovisual Multimedia Services workgroup the question of whether the PS format should also be considered for transmission, given that DVD stores the video data in that format. The solution retained by the Forum is the transcoding of PS into TS packets prior to transmission. Since both formats are very close, the transcoding basically consists on a PS packet parsing.

Since the PS format is not foreseen for transmission over packet networks we only consider in this work TS packets as the data format for transmission of MPEG-2 streams over ATM networks.

6.6.1 Segmentation and Reassembly of Data Units

One of the services offered by the MAAL to the MAAL user is the transmission of fixed length packets to its peer entity. The size of the packets is agreed at connection setup and remains constant during the connection. Therefore, the NAL has to deliver fixed length packets to the MAAL. As such, the segmentation and reassembly function has to provide the following mechanisms to the NAL user:

- delineation of application data packets
- alignment to NAL-PDU boundaries via padding.

In the particular case of MPEG-2, the ADUs are the MPEG-2 system layer TS packets which are of fixed length of 188 bytes. Since this value is a multiple of the size of the MAAL cells, namely 47 octets, there is no need to perform any delineation or padding to align the packets to NAL-SDU boundaries.

The NAL will only have to perform is the assembly of ADUs into NAL-SDUs as a function of the negotiated size.

6.6.2 Selective Data Protection for MPEG-2 Streams

The principle of the PSIP mechanism is to selectively protect the most perceptually relevant information in an audiovisual stream.

In MPEG-2, the most relevant information elements of the bitstream are the headers. These headers can be classified by order of error impact simply by considering the type of information they encapsulate as well as the size. A first priority classification can then be derived:

1. Sequence Header
2. GOP header (optional)
3. Picture header
4. PES header
5. Slice header.

Since three types of predictive pictures exist, it is natural to classify the impact that losses in these pictures have. Intraframe pictures are the most sensitive since errors occurred in these frames will propagate throughout the whole GOP. Errors could therefore last for half a second. The perceptual impact that such losses entail is depicted in Fig. 6.3.

In this experiment, losses have been artificially introduced in each type of frame separately to evaluate their respective perceptual relevance. The perceptual quality is observed through time and the drop and duration of the quality value is then studied.

The cell losses introduced follow a uniform distribution within a single frame. The CLR used has been set to 10^{-4}. Table 6.1 shows both the display and the encoding order of the three damaged pictures.

Picture Type	I	P	B
Encoding Order	12	15	13
Display Order	12	13	15

Table 6.1: *Damaged pictures encoding and display order.*

The figures clearly show that the quality variation depends on the type of picture damaged. The most important impact onto quality is due to a damaged Intracoded frame. From an original average value of 4.2 the quality drops down to 1.5 and the complete recovery is achieved at picture number 36. Note that the quality drops at picture number 10 which is a B picture that uses the next I picture (number 12) which is damaged as a forward reference. Comparatively, the impact that a P picture has onto quality is less important. Even if the quality drops slightly below the curve for the I picture, both curves closely follow each other and in both cases the original quality is recovered at frame 36. Finally, the impact that a damaged B pictures has onto video quality is clearly reduced compared to the precedent two cases in both terms of amplitude and temporal propagation.

It is interesting to note that even if errors do not propagate beyond a GOP, which in this case has been fixed to 12 pictures, the quality is not restored before frame 36 which means two GOPs. This is due to the averaging performed by the MPQM metric.

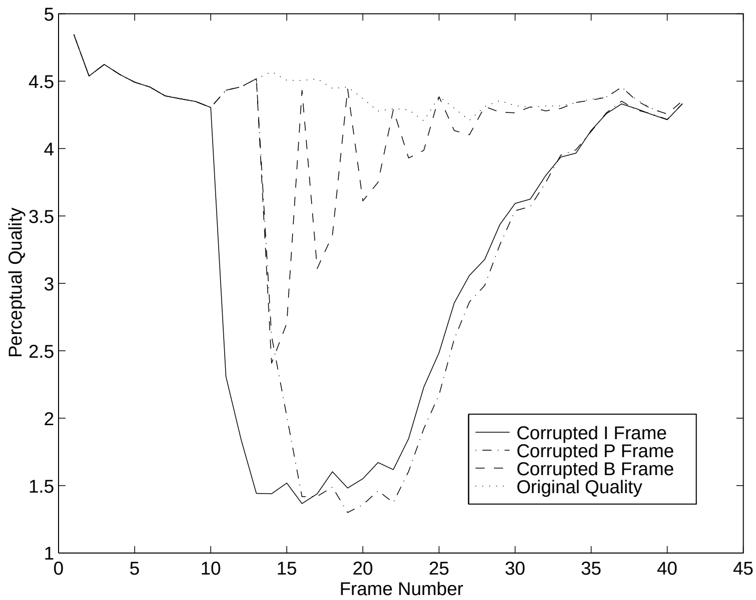


Figure 6.3: *Loss impact onto perceptual quality for I frames.*

From this figures a priority order based on picture types can be derived. The significant quality drop in terms of amplitude and duration observed when an I picture is damaged illustrates the impact that spatial as well as long temporal error propagations have onto quality setting this type of picture as the most important. Intraframe pictures stop error propagation and are used as a reference throughout a GOP. Predictive or P pictures use I frames as references but are also used as references for B pictures. The quality drop is less significant especially in terms of duration and therefore come in second place. The least important being the B or Bidirectional pictures which do not generate any visible temporal propagation which restricts the impact a corrupted B picture has onto the perceptual quality.

Since headers are included into pictures a complete loss impact classification could be done as shown in Tbl. 6.2. If we set priority values for the 3 picture types as well as for the different headers, a hierarchy could be obtained by multiplying the values of the joint picture-header combination. The lowest value being the highest priority.

	Intraframe (1)	Predictive (2)	Bidirectional (3)
Picture Header (1)	1	2	3
PES Header (2)	2	4	6
Slice Header (3)	3	6	9

Table 6.2: *Loss impact classification for MPEG-2 data.*

Outside of this classification are the GOP header, actually optional in MPEG-2, and the Sequence Header which both encompass the pictures. Clearly these have an even higher priority, say priority 0. From this table it is easy to achieve a protection scheme on the basis of priority values. As an example if we want to protect all headers of the I frames and also all picture headers we need a level 3 protection. This does not allow for all combinations of protection but gives a very simple priority-based protection algorithm based on a *priority matrix.*

This type of classification could be further improved by adding the MPEG audio part.

Several experiments done with interactive audiovisual transmission such as the BETEUS teleteaching project [123] have shown that the users are much more sensitive to audio degradation than video. It is therefore natural to consider the audio information as a high priority flow. Since the audio flows require relatively low bandwidth compared to video the protection of the audio flow generates a low overhead.

Also, if it is required it is possible to implement a multi-level PSIP. This means that a level of priority could be assigned a given number of FEC cells. As an example, we could imagine that all picture headers would be protected with a larger number of FEC cells than, say the slice or PES headers, regardless of the type of frame. This will reduce the probability of losing complete frames due to a frame header lost. Other combinations are also possible.

If layered coding is used it is also enviseagable to apply priority levels by layer or by type of picture regardless of the layer. It is actually up to the implementer to develop its own protection algorithm since it is independent from the receiver and can furthermore be downloaded if required.

We propose as a basic protection scheme a PSIP providing the following functions:

- priority table as defined in Tbl. 6.2
- single FEC level (1 cell per PDU).

The situation may occur that multiple headers with different priorities coexist in the same NAL-SDU. If we define $PSIP_h(i,j)$ as the priority for a header of type j in PDU number i, then the protection level is given by the following formula:

$$PSIP_{PDU}(i) = min\{PSIP_h(i,0), \ldots, PSIP_h(i,k), PSIP_{default}(i)\} \tag{6.1}$$

where $PSIP_{PDU}(i)$ is the priority level of PDU number i and $PSIP_{default}$ is the priority selected for PDU number j.

Such a mechanism follows the design principles enumerated in Sec. 6.4. It is simple because it has to identify the relevant data into the NAL-SDUs provides a low overhead since it selectively protects the headers and it does not require a long processing delay.

6.7 Simulation Experiments

This set of simulations aims at evaluating the efficiency of the PSIP mechanism compared to an equivalent transmission over AAL5. Since PSIP produces overhead, two different quantizer scales have been considered, 26 and 28 which both produce a CATV quality, in such a way that 10% more bandwidth in average is required by the former. The saved bandwidth obtained by the lower quality encoding will be used for data protection (PSIP). This will therefore lead to an equivalent mean, but not necessarily peak, cell rate because FEC cells may be added in the high activity periods of the lower quality stream. Still, the difference is negligible as shown in Tbl. 6.3 which presents the actual traffic parameters achieved for both streams. The simulation setup used is the same as described in Chap.5.

mquant	Peak Rate (Mbps)	Mean Rate (Mbps)
26	7.36	4.19
28	6.56	3.64

Table 6.3: *Traffic parameters for the two bitstreams.*

6.7.1 Network Performance

The network performance results are shown in Fig. 6.4. The CLR figures observed look similar than in earlier experiments. Even if in this experiment both bitstreams do not have exactly the same cell loss process, since the streams are not identical, the CLR figures still show the better performance of the MAAL. The CLRs achieved by the MAAL are close to an order of magnitude smaller than those for AAL5.

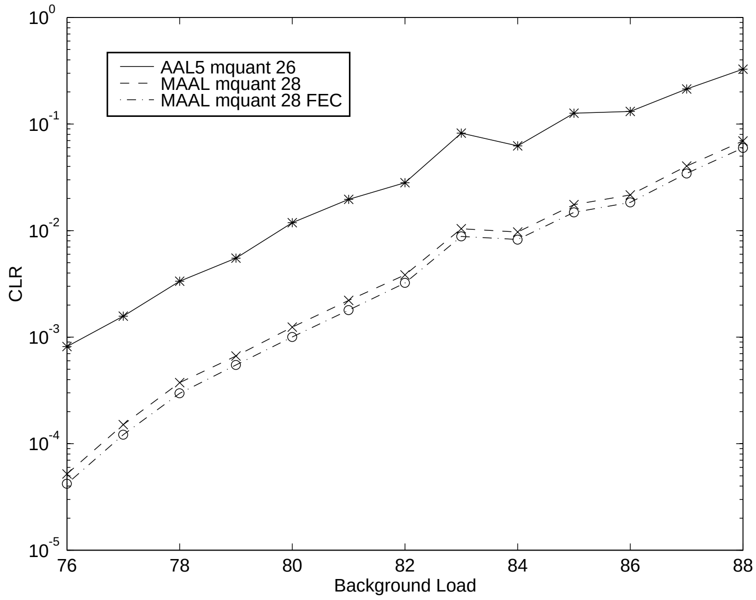


Figure 6.4: *Cell loss ratio performance of PSIP.*

The dashed MAAL curve shows the total CLR observed. The dot-dashed line takes into account all the cells recovered by the FEC mechanism. Obviously the second curve is below which indicates that some lost cells have been recovered. From a network QOS point of view, the improvement is not significant because only 10% of the cells are protected and may be recovered.

The protection mechanism being selective, the CLR metric is not adequate to measure the efficiency of the PSIP. To evaluate the improvement achieved by the PSIP mechanism, we define the FEC Gain G_{FEC} as follows:

$$G_{FEC} = \frac{Recovered_{FEC}}{Loss_{FEC}}, \tag{6.2}$$

where $Loss_{FEC}$ denotes the number of unrecovered protected cells and $Recovered_{FEC}$ the number of *recovered* protected cells. This Gain is in fact the ratio between recovered and unrecovered cells protected by the PSIP mechanism.

As shown in Fig. 6.5 the FEC Gain achieved decreases with the increasing CLR. The higher the CLR the higher the probability of observing consecutive cells lost in a single protected packet which tends to decrease the recovery capacity, thus the FEC gain. This means that under high CLR values, a single FEC cell does not achieve enough protection. Beyond 10^{-3} the gain drops below 1 which means that less than 50% of the protected cells are recovered. Even, for a CLR of 10^{-2} almost 40% of the lost cells protected by PSIP are recovered which gives evidence of the good performance achieved by the protection mechanism.

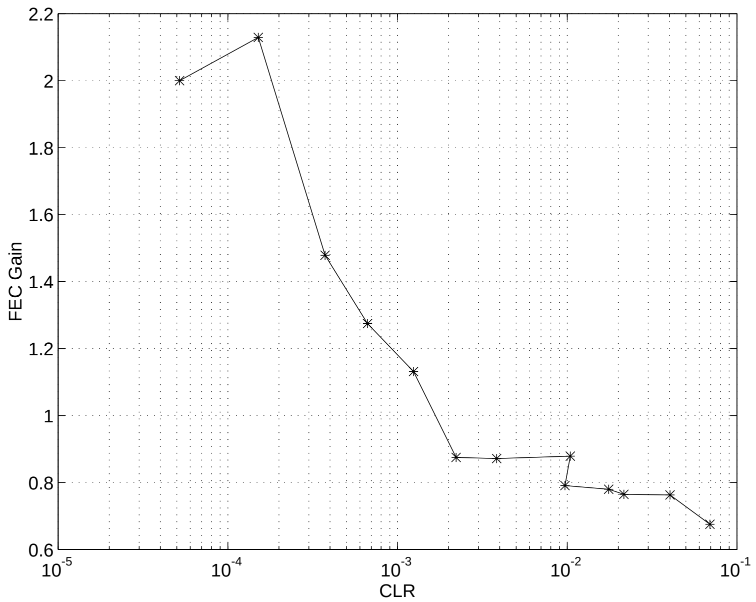


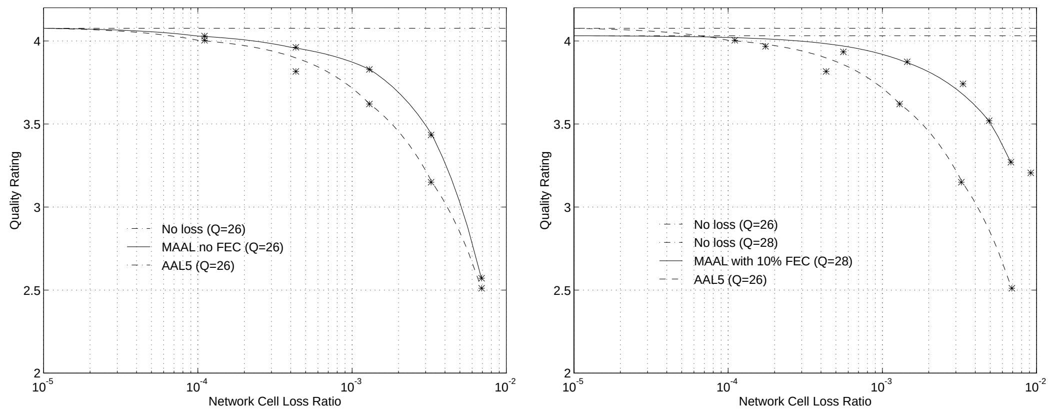
Figure 6.5: *FEC gain.*

The PSIP selectively protects a small fraction of the data. The network results show that some recovery is achieved. However, all the interest of the PSIP mechanism resides in the perceptual importance of the recovered cells. The impact that such a small fraction of recovered cells may achieve onto the user perception is the subject of the next paragraph.

6.7.2 User Perceived Performance

Figure 6.6 shows the perceptual quality assessment using the NVFM metric as a function of the CLR for the transmission of MPEG-2 VBR streams over AAL5 and MAAL. Fig. 6.6 (a) compares the quality of the same video stream transmitted with both AALs. The MAAL shows a better behavior under medium to high CLRs without PSIP. Note that both the AAL5 and MAAL figures begin to diverge for a CLR value close to 10^{-5}. This is due to the better performance of MAAL. However, under very high CLRs both figures converge again. This is due to the limited burstiness of the cell loss process which groups the losses and therefore reduces the negative effect that the AAL5 packet discard mechanism produces. However, the MAAL is still more reliable without any extra overhead which still can be explained by the fact that the cell loss clustering effect due to the correlated traffic is actually very limited which clearly penalizes the AAL5.

Figure 6.6 (b) depicts the results obtained for the transmission of both an unprotected video stream over AAL5 and a PSIP protected one over MAAL. The figure which gives the best quality is the transmission of the lowest video encoding quality stream with MAAL and PSIP. It is interesting to note that even if the quality of the encoded stream is lower, the quality at the receiver is better for a CLR beyond 10^{-4}. For higher CLR values both figures diverge increasing the difference in quality achieved by the MAAL. Even if we have shown that the burstiness of the cell loss process is limited it is true that the FEC mechanism is penalized under such conditions. Bearing in mind that the protection mechanism adds a single redundancy cell, configuration which is not optimized for such bursty loss processes, we can conclude that further improvements on the PSIP mechanism could lead to better results. Consequently, a tradeoff may be found which optimizes the perceived quality (the



(a) No FEC

(b) FEC

Figure 6.6: *Quality rating vs. network CLR for VBR video transmission over AAL5 and MAAL.*

bitrate) and the protection efficiency (the overhead).

Figure 6.7 shows the perceptual quality assessment as a function of the CLR using the SHAT metric introduced in Chap. 3. The results are close in shape to the ones obtained with the MPQM metric. The quality remains nearly constant up to a certain value beyond which the quality drops at a fast pace. The curves also show that beyond 10^{-3} the quality of the AAL5 streams transmitted over AAL5 drop below the quality of the MAAL streams. The difference in quality increases with the CLR. The major difference between these results and the ones achieved by the MPQM metric resides in the variations in quality captured by the SHAT metric. While the MPQM values drop down to 2.5 for CLR values close to 10^{-2}, the SHAT values do not exceed 4.5. This is a well known problem of the SHAT metric which has been optimized for MPEG-1 streams and does not behave consistently in the low bit rate range of MPEG-2.

Still, the use of a different perceptual quality metric, the SHAT, shows that the proposed AAL and NAL layers deliver better video quality than a standard transmission over AAL5 under cell loss without any overhead in terms of bitrate.

6.8 Statistical Multiplexing Gain

The results of Sec. 5.3.2.2 showed that under the studied conditions, significant potential for statistical multiplexing gain could be expected due to the low correlations observed in the cell loss processes of VBR video streams.

The results observed in the precedent section show that under relatively high CLRs, in the range of $10^{-4} - 10^{-3}$, the perceptual quality remains very close to the original encoded quality. These results suggest that it is possible to fix relatively high CLRs and still achieve a high quality video.

To evaluate the potential gain offered by the proposed layers in terms of SMG we calculate the maximum link utilization as well as the maximum number of connections allowed for a video stream as functions of the CLR.

The approximation used introduced in [168] assumes the traffic to be conforming to two

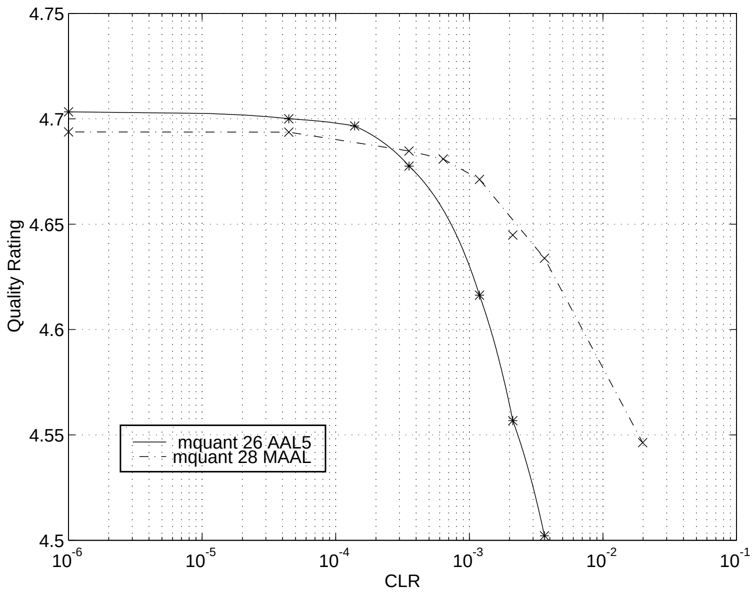


Figure 6.7: *SHAT quality rating vs. network CLR for VBR transmission over AAL5 and MAAL.*

connection parameters, the PCR and SCR. The model is a worst case approximation of a traffic conforming to a dual leaky-bucket whose MBS does not influence the CLR. The sources are assumed to be of On-Off type. Then it is possible to derive the maximum number of connections for a given CLR:

$$N(p) = sup_{N\geq 0}\{N, CLR(N,C,r,p) < \epsilon\}, \tag{6.3}$$

and the CLR is given by:

$$CLR(N,C,r,p) = \frac{\sum_{n=[\frac{C}{p}]}^{N}(np-C)\begin{pmatrix} N \\ n \end{pmatrix}\alpha^n(1-\alpha^{N-n})}{Nr}, \tag{6.4}$$

where C is the link capacity, r is the SCR p the PCR and ϵ the negotiated CLR.

Knowing the relationship between the maximum number of connections and the allowed CLR, it is possible to calculate the SMG by calculating the ratio between $N(p)$ and $N(PCR)$ which is the maximum number of connections allowed by peak cell rate allocation and is expressed as:

$$N(PCR) = \frac{C}{p}. \tag{6.5}$$

Thus, the SMG is given by:

$$SMG = \frac{N(p)p}{C}. \tag{6.6}$$

Figure 6.8 depicts the SMG as a function of the CLR for the video stream encoded with a quantization factor of 26 (see Tbl. 6.3 for details) and two link capacities of 155 and 622 Mbps.

Both curves show the interest of using statistical multiplexing. If cell losses can be tolerated by the data then important gains can be achieved. The perceptual quality results depicted in Fig. 6.6 show that for CLR values up to 10^{-4} the quality remains close to the original. Then it is possible to allow 18% and 37% more connections depending on the link rate. If the PSIP mechanism is used it is possible to push the CLR limit to almost 10^{-3} which allows to increase the number of connections to 24% and 42% respectively. Therefore it is economically interesting to use VBR sources because they allow statistical multiplexing. In addition, the combination of the proposed protocol layers allows to increase the video tolerance to loss.

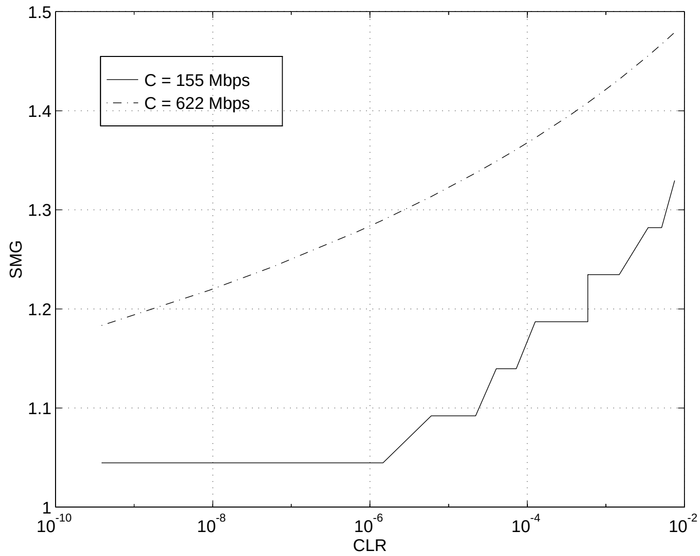


Figure 6.8: *Statistical multiplexing gain vs. CLR for a VBR video stream encoded open-loop with a quantization factor of 26.*

6.9 Conclusion

We have introduced in this chapter a network adaptation layer for interactive multimedia communications. This layer provides an intelligent and minimum overhead protection of video information. The PSIP takes advantage of the hierarchical organization of the compressed video data, common to the majority of compression algorithms, to perform a selective data protection based on the perceptual relevance of the syntactic information. This, however, requires the NAL to be codec-specific in order to identify the relevant data structures within the bitstream.

Based on a set of design principles, we have designed and tested an MPEG-2 specific NAL. The experiments show a significant improvement in terms of perceptual quality compared to equivalent transmissions over AAL5.

The perceptual quality improvements allow for a better network utilization through statistical multiplexing. The experiments show that for CLR values in the range $10^{-4} - 10^{-3}$, the gains achieved for a CATV quality video stream in a 622 Mbps link are in the range of 37 to 42% leading to a significant cost reduction for such interactive services. This brings to the fore the interest of using VBR sources and the new protocol layers which are able to increase the tolerance to loss of compressed video streams. Such improvement is achieved by using a selective open-loop protection mechanism which adds very low delay and overhead and can therefore be used for interactive services in point-to-point as well as multipoint configurations.

The significant gains achieved by the usage of the combined NAL-MAAL layers is explained by the complementarity of the functions provided by each of the layers. If the NAL had to be used on top of other AALs such as AAL5, then potential gains are still to be expected. However, the functions performed by the MAAL, in particular the generation of FEC data should be moved to the NAL. In that case as shown in Sec. 4.5.4, the data protection would become packet-based, therefore losing part of its recovery efficiency.

Chapter 7

A Prototype Software Implementation of the Multimedia ATM Adaptation Layer

7.1 Introduction

We describe in this chapter a proof-of-concept prototype software implementation and the related experimental results of the multimedia ATM adaptation protocol layer designed and tested by simulation in the precedent chapters. The MAAL prototype is mainly based on the specifications described in Appendix A. However, due to the choice of a user space implementation, justified by a simple and fast development and debugging cycle, some modifications have been incorporated to improve the software speed.

The interest of such implementation is twofold. First it allows to confront the simulation results with experimental data. Second it is possible to verify if the design is consistent and allows for a real system implementation.

The main characteristics of the FORE's ATM boards are described in the next section. Such characteristics have forced some design choices also described. The choices made for the layer implementation are described in Sec. 7.3. Section 7.4 presents the testbed and the measurement results obtained.

7.2 The FORE Systems ATM Boards

To develop an AAL on a FORE ATM board, two possibilities can be envisaged; the first consists on writing a driver level code for the FORE board. The implementation is more efficient but this method has several drawbacks for a software prototype. To test driver code, it has to be downloaded in the board firmware. It is more difficult to debug and each modification needs to repeat downloading. The second possibility is to write the AAL in a standard high level programming language on top of AAL0. This AAL, available on FORE boards as shown in Fig. 7.1, is accessible via the FORE Application Programming Interface (API) which provides C code interface to the board management and communication routines. The AAL0 is transparent and gives direct access to the ATM layer making it ideal for such prototype implementations. It lets users access the ATM cell headers (see Sec. 7.3.1), but VPI/VCI, HEC and PTI values are handled by FORE's routines and are not available to the user. In this second alternative, the code is developed in user space which simplifies the compiling and debugging process.

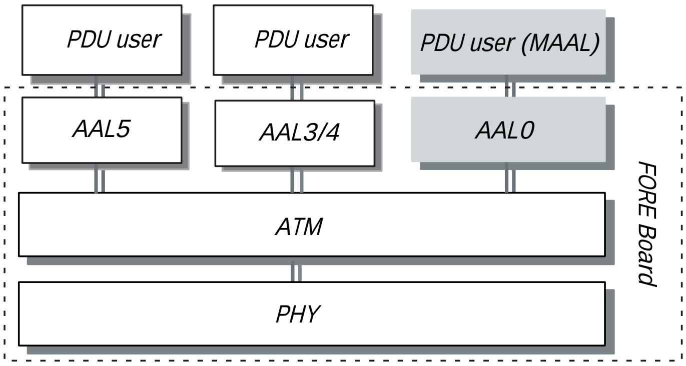


Figure 7.1: *Fore ATM board AALs.*

The FORE API provides the interface to a connection-oriented mechanism based on a client-server model. The client sends data while the server can receive data from different incoming connections.

The API routines work on multiple platforms. The routine library is based on UNIX streams and communicates with the ATM manager via the Data Link Provider Interface.

The data transfer is done in a best effort way in the sense that at connection setup a target bit rate is given to the client but is not guaranteed. This is partly due to the fact that the system performance depends heavily on the workstation load. In addition, any flow control and retransmission is left as a user implementation option.

The signaling used in these experiments is provided by a proprietary protocol called SPANS. The current version of the ATM switch software allows also to use UNI 3.1 signaling.

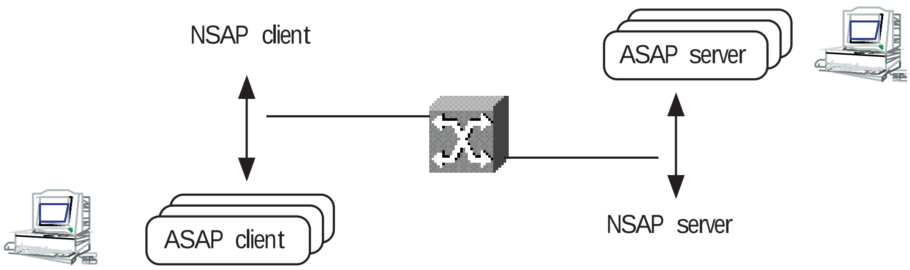


Figure 7.2: *NSAP/ASAP illustration.*

Clients and servers are identified by a unique Application Service Access Point (ASAP). If connected to a switch, the terminals must also have a Network SAP (NSAP) as illustrated by Fig. 7.2. A terminal can receive only a single connection to a given ASAP as shown in Fig. 7.3. To establish a connection, it is mandatory to know both the NSAP and the ASAP. FORE provides a routine *atm_gethostbyname* that allows one to know the NSAP via the workstation's name.

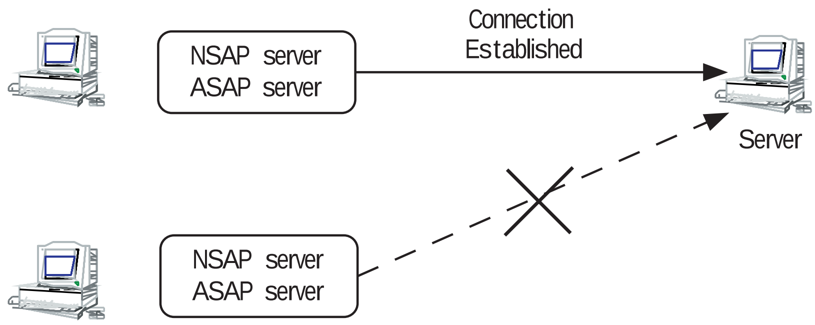


Figure 7.3: *Single application service access point.*

7.3 The MAAL Implementation

7.3.1 MAAL Classes

FORE APIs provide a C language interface. To be consistent with this API, the implementation has been done in C++. It has the advantage of providing data protection. C++ makes also use of the inheritance property. Since FORE API works on multiple platforms, the current MAAL version runs on HP VUE and SUN Solaris workstations.

The core part of the MAAL software consists of three C++ classes:

- server: groups the server (receiver) specific functions. It can receive data from multiple incoming connections
- client: groups the client (sender) specific functions. May perform multiplexing
- common: groups the functions common to both client and server classes.

The main part of the code is contained in the common class. The main difference between the other two classes is the connection setup code. Both client and server classes inherit from the common class avoiding redundancy in the code. This gives more flexibility since modifying the common class automatically modifies the other two classes by inheritance. The server and client classes contain specific functions not covered in the common class. The C++ functions and interfaces are described in detail in [169] and in appendix C.

7.3.1.1 The Common Class

The functions provided by the three classes follow the same philosophy as FORE's API routines. This gives a consistent interface to the user.

All the functions available to the user (public functions) return an error code depending on the type of event: a negative code is returned in case of error and a positive one in case of success. Table 7.1 summarizes the public functions of the common class.

The common class also contains some private functions used for its internal management. These functions have been dissociated from the public ones to improve flexibility. For instance, the FEC algorithms have been implemented as private functions. If a different algorithm has to be implemented only two functions require modification. Table 7.2 summarizes the private functions of the common class.

Function Name	Description
open_connection	opens a MAAL connection with the requested QoS. Returns a connection identifier for each call. The PDU size has to be indicated.
close_connection	closes a MAAL connection.
send_PDU	sends a PDU through the specified connection. The PDU has to be non-empty and must have the size specified at connection setup.
recv_PDU	receives a PDU through the specified connection. Returns the number of received octets. The PDU has to be pre-allocated.
get_params	allows to obtain a copy of the parameters from a specific connection.
enable_FEC	enables calculation and transmission of FEC cells.
disable_FEC	disables calculation and transmission of FEC cells.
maal_error	returns an error message corresponding to the current error.

Table 7.1: *Public functions of the MAAL common class.*

Function Name	Description
calculate_FEC	calculates the FEC cells for a given PDU.
decode_FEC	decodes errored PDUs with the received data and FEC cells.
decode_PDU	allows to decode the segments of a received PDU. In case of cell loss it calls decode_FEC. When no recovery is possible it inserts dummy cells to guarantee PDU size integrity.
encode_PDU	performs the segmentation of PDUs into MAAL-SDUs inserts the MAAL header and the FEC cells when requested.

Table 7.2: *Private functions of the MAAL common class.*

7.3.1.2 The Server Class

The public functions of the server class are all identical to the common class public functions but the *open_connection* function. In the current implementation, two schemes are available to receive the ATM cells. They ma ybe selected by the user at connection setup. This choice is justified by the fact that each method has some advantages compared to each other.

The simplest method is based on FORE's *atm_recv_null* routine. The advantage of this method is that it is simple and has low dependencies on external factors to have a good reception. The receiver's UNIX pipe has a size of 8 kbytes. The drawback is that if the pipe gets full (parent process too slow) data may be lost.

The alternative method uses a child process principle such as the one depicted in Fig. 7.4. The child process manages only the reception of ATM cells. Once cells are received, they are sent to the parent process via a standard UNIX pipe. Therefore, the parent gets the cells from the pipe and not directly from FORE's routine.The advantage is that the pipe acts as a small buffer allowing more flexibility in terms of reception process speed. Moreover, it is possible to define a batch size of cells with *atm_setbatchsize* that the *atm_recv_null* will store prior to passing the data to the parent process. This reduces the size of the UNIX buffers and increases the software speed, since the FORE routine does not pass the data cell by cell but in batches.

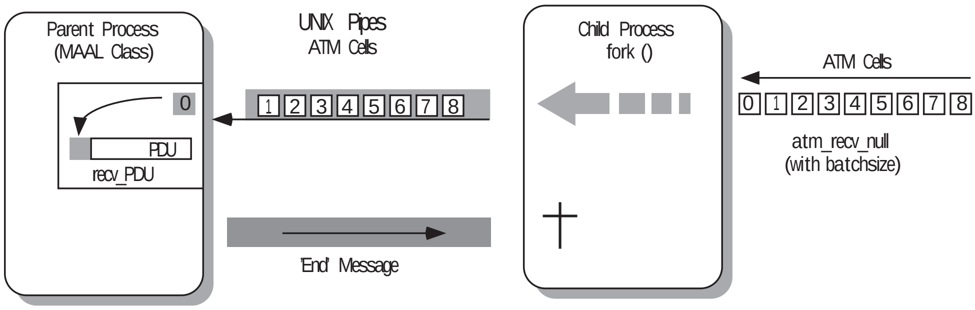


Figure 7.4: *Child process implementation.*

One of the problems of FORE's API is that it does not allow to work with a concurrent server model directly with the provided C routines. To overcome this issue, a free software package nbatm(non-blocking ATM) has been modified and integrated into the MAAL software. The advantage of nbatm is that it allows to have a single listening process in the standard ASAP. This leaves the standard ASAP free to listen for other connections. Once a connection request arrives directed to this access point, a new listening process with a random ASAP is created and the request is rerouted to that process. The client gets connected to the random ASAP for the rest of the communication. The sequence of events involved in the connection setup process depicted in Fig. 7.5 is as follows:

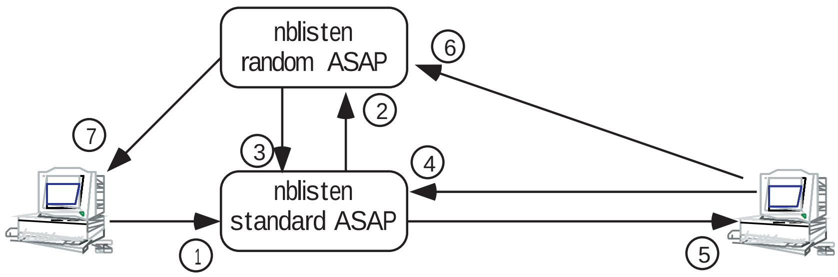


Figure 7.5: *Connection using the nbatm library.*

1. the terminal in listen mode launches the listen process in the standard ASAP
2. the listen process launches a new listen process in a random ASAP
3. The random ASAP is communicated to the listen process in the standard ASAP
4. a client request arrives to the standard ASAP
5. a connection is established to communicated the new ASAP to the client. The connection is released
6. The client connects to the random ASAP for the rest of the connection.

7.3.1.3 The Client Class

The public functions of the client class are all identical to its parent the common class.

One of the main problems that the FORE software has is the allocation of small UNIX buffers. These buffers are required by the *atm_recv_null* routine to collect the cells prior to passing them to the higher layers. Due to the real-time constraints of the MAAL, very small buffers are required (e.g. 4 or 8 cells). Due to unknown reasons, the buffers may not be allocated. FORE support was unable to solve the problem. The proposed solution has already been described in Sec. 7.3.1.2. However, if the incoming cell rate is highly variable, the allocation problem worsens leading to large consecutive cell losses. Since the client cell rate heavily depends on the workstation load a rate control has been implemented to overcome the problem.

The proposed rate control is based on a moving average principle depicted in Fig. 7.6. The mean time required to send 4000 PDUs is measured and updated for every new PDU sent. The average time is then compared to the theoretical time based on the target bit rate specified by the user. According to the value of the error e, the bit rate is changed to reach the target value following one of two implementations:

$$
\begin{aligned}
e &= [\Delta_{4000real} - \Delta_{4000target}] \\
T_{upd} &= T_{target} \pm e^2 \\
T_{upd} &= T_{target} \pm 100 \times e,
\end{aligned}
$$

where $\Delta_{4000real}$ is the moving average time, $\Delta_{4000target}$ is the target time interval and T_{upd} is the new target time interval from which the bit rate is calculated.

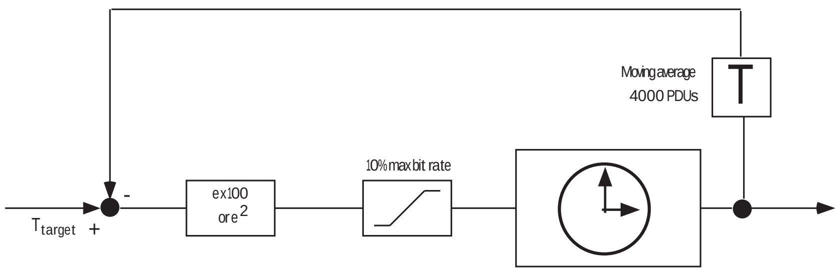


Figure 7.6: *Cell rate control principle.*

7.3.2 The Forward Error Correction Implementation

The forward error correction mechanism implemented is based on the RSE codes principle described in Chapter 4. However, due to software performance problems, a simplified version has been chosen for this implementation.

The algorithm is based on *XOR* operations. The FEC cells are generated by XORing the data cells. The advantage of this method is its simplicity and speed. The drawback is that when several FEC cells are required, the XOR operation does not generate multiple linearly independent redundancy cells.

The implemented technique finds its rationale on the simulation results of the precedent chapters which showed the small correlation of the cell loss process.

Figure 7.7 illustrates the principle implemented in the MAAL. In this example, the FEC cells are generated as follows:

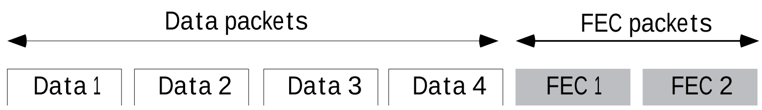


Figure 7.7: *FEC implementation example.*

- FEC1 = DATA1 XOR DATA3
- FEC2 = DATA2 XOR DATA4.

This mechanism does not generate linearly independent FEC cells which means that not all combinations of cell losses may be recovered. Some of the recoverable cases of two cell losses are:

- DATA1 and DATA2
- DATA2 and DATA3
- DATA3 and DATA4
- DATA4 and FEC1.

However, the loss of DATA1 and DATA3 cannot be recovered with this technique. The redundancy generated by this method compared to the recovery ratio it offers is far from being optimal, but allows a very simple and fast implementation which is the trade-off that has been considered in our case. The implementation of a full FEC able to generate multiple redundancy cells requires complex algorithms as described in Sec. 3.3.2.1. Albeit fast software implementations can be found, their integration within the MAAL posed problems in particular related to operation speed.

The generalization of this technique to PDUs constituted of n cells is as follows:

- n is even. The cell FECi is calculated by XORing cell number i and cell number $\left[\frac{n}{2}+i\right]$ as shown in the table:

	Data Cell Number	FEC Cell
1	$\frac{n}{2}+1$	FEC1 = 1 XOR $\frac{n}{2}+1$
2	$\frac{n}{2}+2$	FEC2 = 2 XOR $\frac{n}{2}+2$
	$\vdots$	
i	$\frac{n}{2}+i$	FECi = i XOR $\frac{n}{2}+i$
	$\vdots$	
$\frac{n}{2}$	n	FECn = $\frac{n}{2}$ XOR n

Table 7.3: *Generation of FEC cells for n even.*

- n is odd. The algorithms differs from the precedent for the calculation of FEC1. This cell is calculated with cell number n and cell number 1. In addition, the emission order of FEC1 and FECn has been inverted to allow the recovery of the last data cell in any case.

Data Cell Number			FEC Cell
1	$\frac{n-1}{2}+1$	n	FECn = $\frac{n-1}{2}$ XOR n-1
2	$\frac{n-1}{2}+2$		FEC2 = 2 XOR $\frac{n-1}{2}+2$
	$\vdots$		
i	$\frac{n-1}{2}+i$		FECi = i XOR $\frac{n-1}{2}+i$
	$\vdots$		
$\frac{n-1}{2}$	n-1		FEC1 = 1 XOR $\frac{n-1}{2}$ XOR n

Table 7.4: *Generation of FEC cells for n odd.*

7.4 Experiments

7.4.1 System Configuration

The experiments have been carried out in the laboratory's ATM LAN depicted in Fig. 7.8. The LAN is composed of several workstations equipped with FORE SBA-200 ATM boards connected to a FORE ASX 200 ATM switch. An HP 75000 series ATM test equipment is also connected to the ATM switch allowing the generation of background traffic and also to perform ATM layer measurements. The links between the workstations and the switch are 155 Mbit/s links.

Both HP 700 and SUN SPARC workstations are connected to the ATM LAN via FORE ATM boards. The MAAL implementation has been successfully compiled and tested in both types of workstations.

The experiments have mainly consisted in file transfers. The lack of real-time MPEG-2 encoders and decoders did not allow us to perform full real experiments. Therefore, the encoding and decoding processes were performed offline. This however, does not remove any generality to the results since the network performance parameters collected do not depend on the end systems.

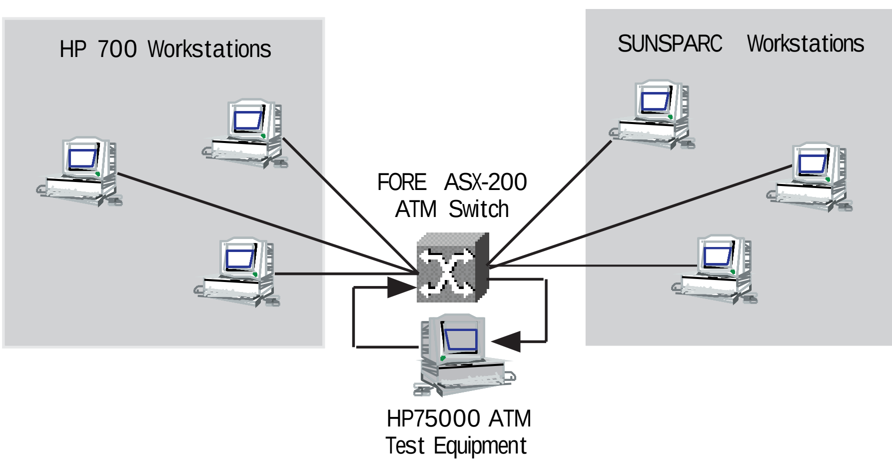


Figure 7.8: *Experimental ATM LAN configuration.*

Number of FEC cells	Background load (%)	
	84%	96%
1	8.95 μs	9.21 μs
2	9.31 μs	9.73 μs
3	9.45 μs	10.13 μs
4	10.68 μs	10.93 μs

Table 7.5: *Average error detection processing time as a function of the background load.*

7.4.2 MAAL Software Prototype Performance

7.4.2.1 MAAL Data Rate

One of the main factors that influenced the performance of the software prototype was the BATCHSIZE constant. As explained in Sec. 7.3.1.2 this constant defines the size of the UNIX buffers used to absorb the bit rate variations between the parent and child processes. Several tests have shown that setting the variable to a small number improves the software performance. The other major element found to influence the software performance was the workstation load. Without the possibility to prioritize processes it has not been possible to guarantee that only the real-time processes were running.

Considering the aforementioned issues, tests were carried out in time frames with low utilization of the workstations. The current MAAL software implementation has been able to transfer data at values of *up to 6 Mbit/s*. When cell losses begin to appear, the bit rates slow down to *5 Mbit/s* due to the extra processing required to correct the data.

7.4.2.2 Error Correction Implementation Performance

To evaluate the performance of the FEC implementation, the processing times required to correct lost cells have been measured. The measurements have been carried over both the CBR and Mixed background traffic experiments. Table 7.5 summarizes the average processing times measured. The processing times measured in the ATM boards have a small number of large fluctuation that in fact increase the average values to figures not really representative of the majority of the processing times. In the particular case of the FEC recovery, these large variations could reach times of up to 1 ms and the general case was near the μs.

The figures show that the processing time increases with the number of FEC cells linearly. However, the processing time required by a 4 FEC cell stream is not twice the time required by a 2 FEC cell stream. The explanation to this phenomenon could be that when no losses occur, which is the general case, the reassembly times are almost the same for all cases which then reduces the average time required to correct cell losses.

7.4.3 Multimedia AAL Network QoS Performance

All experiments have been performed with CBR foreground traffic only. Due to the bandwidth limitations of the prototype and the variable processing time problems observed experiments with VBR video traffic were not performed.

Two sets of experiments have been carried out. The first uses CBR only background traffic. The second uses a mix of CBR and Bernoulli traffic to achieve a global VBR background traffic. The test equipment does not allow to generate complex traffic patterns. Therefore, Bernoulli traffic was chosen as a VBR background traffic profile.

7.4.3.1 Performance with CBR background Traffic

Figure 7.9 shows the percentage of received PDUs as a function of the load for five different overhead cases. The first observation is the big difference noticed between the unprotected and the protected case. The measurements have been performed upon reception of the PDUs without any particular check of the contents. The four FEC cases show a consistent behavior. The number of received PDUs remains close to 100% and then drops with the increasing load. The unprotected case behaves more erratically. Such a high percentage of lost packets can hardly be explained by the network load. The traffic profiles, both foreground and background are CBR and thus deterministically allocated. The reason to this high packet loss is to be found in the receiver itself. When increasing the load, the receiver cannot cope with all the incoming packets. This generates timeouts and buffer overflows which causes the observed losses. This buffer allocation problem is one of the main issues that have difficultly been solved in this implementation. The variable processing times observed in the FORE boards generate timeouts and buffer overflows. The reason to this highly variable processing times is not known.

The conclusion related to Fig. 7.9 is that the FEC streams are more reliable than the unprotected stream even if the bit rate is higher which should tend to increase congestion. The efficiency increases with the number of FEC cells appended.

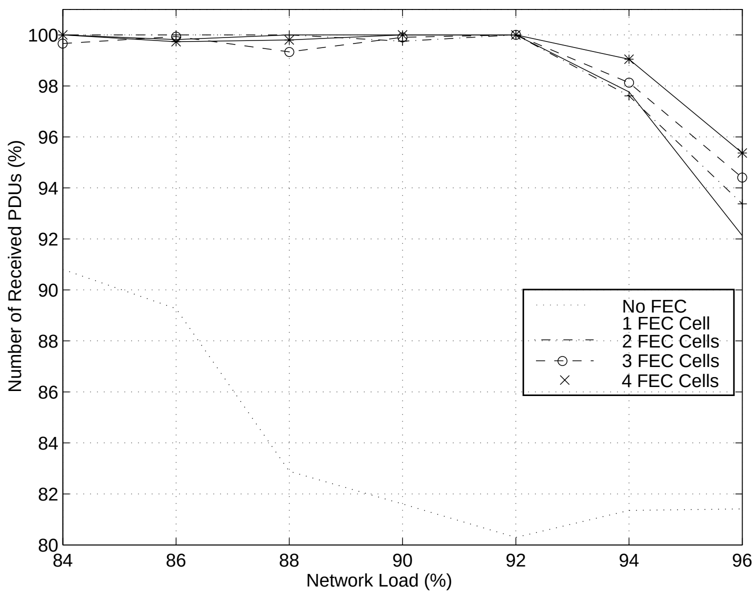


Figure 7.9: *Percentage of received PDUs as a function of the background load.*

Figure 7.10 depicts the number of *fully lost* PDUs. This means that none of the cells forming a PDU have been received. It is necessary that eight consecutive cells belonging to the same PDU have been lost for this to occur which is a relatively rare event as shown by the small percentages measured and the results of Chap. 5. The curves show some influence due to the FEC overhead. The bigger the overhead, the higher the loss ratios measured. This situation changes for loads beyond 92% where the unprotected stream curve starts to increase with the highest slope. For loads over 94% the unprotected stream suffers from the highest PDU loss ratio. This means that for such high loads the congestion becomes important. Since no correction is possible for the unprotected stream then packet losses are not negligible anymore.

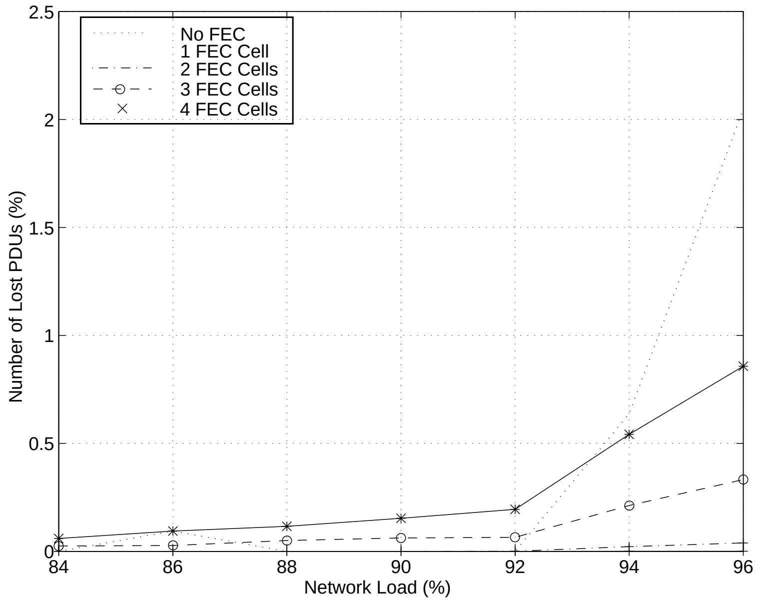
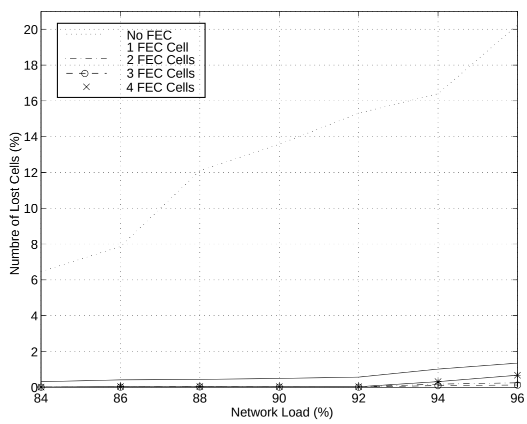
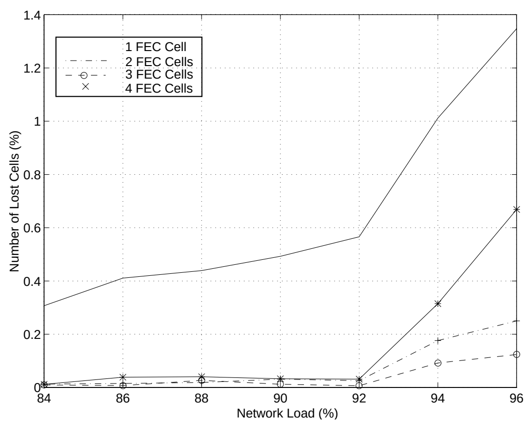


Figure 7.10: *Percentage of lost PDUs as a function of the background load.*

Figure 7.11 depicts the loss process observed at the cell level. Figure 7.11 (a) shows the CLRs for all the streams. Again, the unprotected stream exhibits the worst performance. The loss ratios measured are more than an order of magnitude bigger.



(a) All Cases



(b) FEC Cases Only

Figure 7.11: *Percentage of lost cells as a function of the background load.*

Figure 7.11 (b) shows a detail of figure (a). Only the protected streams are plotted which allows to see in detail the performance of the four protected streams. This figure shows that using 1 or 4 FEC cells is not optimal. The recovery efficiency achieved with a single FEC cell is not enough to cope with all cell losses. On the other side, 4 cells are enough to take care of almost any losses but it adds an overhead that contributes to the congestion and thus does not achieve the best performance of the four cases. The recovery performance obtained when using 2 and 3 FEC cells shows that these values are optimum for the case studied here. Using 3 cells achieves the better performance of all the cases. However, if these results are confronted with those of Fig. 7.9 the optimum number of FEC cells seems to be 2 because it performs better in terms of full PDUs lost and gives very close results in terms of cell losses.

Figure 7.12 depicts the PSDs of the sampled cell loss processes. All the spectra are flat which characterizes an uncorrelated loss process and explains the improvements achieved by the FEC mechanism. The PSD figures are similar to the ones obtained via simulation in Sec. 5.3.1. The main difference is that in the experimental case, there is no evidence of a periodic component which clearly appeared in the simulated case. In fact, the real system is not slotted such as the simulated one. Therefore, the CBR stream is not exactly constant. Consequently, the sampled times are not multiples of the same cell emission interval, thus explaining the absence of peaks in the spectra.

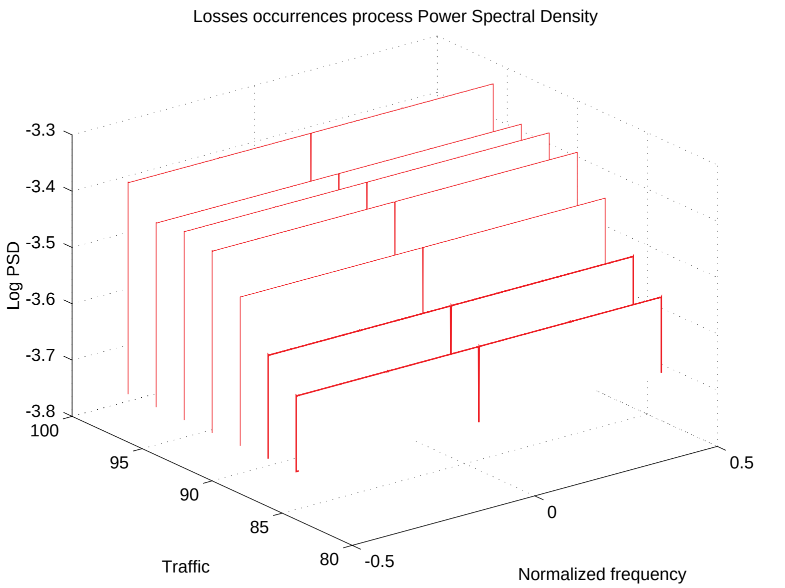


Figure 7.12: *Cell loss process PSD. CBR background traffic and 2 FEC cells.*

7.4.3.2 Performance with Mixed background Traffic

The traffic used in this set of experiments is a superposition of bursty traffic described in Sec. 7.4.1 with a CBR traffic. The mix generates an overall VBR background traffic. To create different loads, the CBR traffic load has been increased in steps of 2%.

The utilization of a bursty background traffic modifies the performances of the recovery mechanisms. As shown in Fig. 7.14 using 3 and 4 FEC cells leads to a relatively poor performance, even worse than for an unprotected cell stream. In this case, the optimal performance is achieved for 1 and 2 FEC cells that achieve almost 100% recovery in all tested cases. In Fig. 7.11 the optimum was achieved for 2 and 3 cells. Therefore 2 FEC cells seems to be the best compromise for all traffic configurations tested.

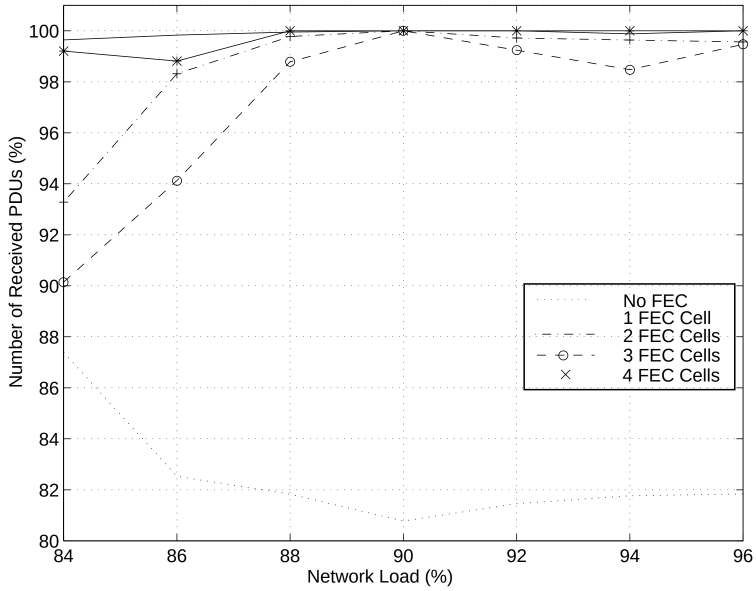


Figure 7.13: *Percentage of received PDUs as a function of the background load.*

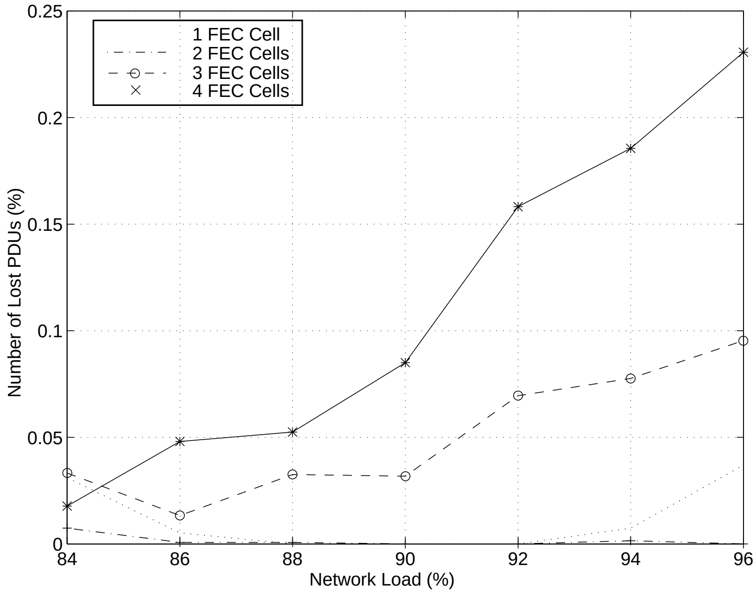
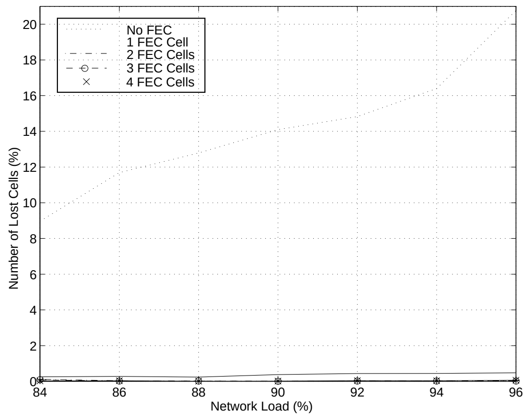


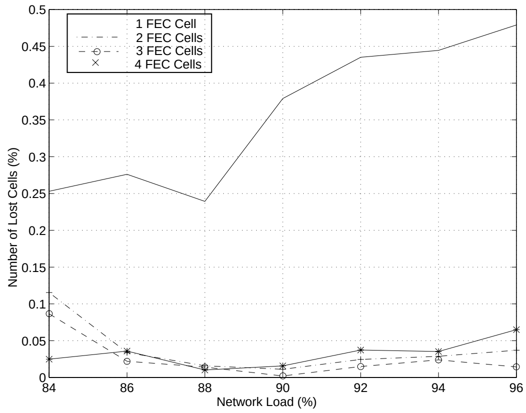
Figure 7.14: *Percentage of lost PDUs as a function of the background load.*

Figure 7.15 (a) shows that the unprotected cell stream suffers from higher CLR than the protected streams. This seems to be in contradiction with the precedent figure showing that the use of 3 and 4 FEC cells achieve much higher PLRs. Indeed, the situation shows that increasing the overhead improves at a cell level the loss ratios observed at the receiver. However, the extra traffic due to this overhead groups the losses thus generating more packet losses. This cell loss clumping does not occur at the network level. The extra processing time due to supplementary losses and to extra FEC cells generates timeouts at the receiver which generates such grouped losses.

Figure 7.15 (b) leads to the same conclusions than Fig. 7.11 (b). The performances achieved with 1 and 4 FEC cells are not as good as using 2 or 3 redundancy cells. Taking into account the results of Fig. 7.14, 2 overhead cells achieves the best performance.



(a) All Cases



(b) FEC Cases Only

Figure 7.15: *Percentage of lost cells as a function of the background load.*

Figure 7.16 which depicts the PSD of the cell loss process also shows the same behavior observed in Fig. 7.12 and in the simulations of Chap. 5. The cell loss processes appear to be uncorrelated explaining the good loss recovery performance achieved by the FEC mechanism and validating the low traffic source assumption used in Chap. 4 for the design of the MAAL.

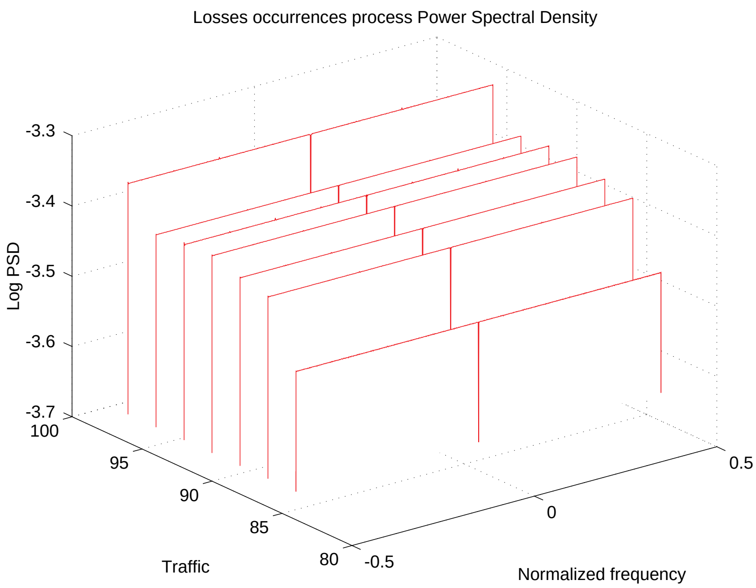


Figure 7.16: *Cell loss process PSD. Mixed background traffic and 2 FEC cells.*

7.5 Conclusions

We have described in this chapter a proof-of-concept prototype software implementation of the multimedia AAL. The prototype layer includes a simplified forward error correction mechanism. The tests have shown the layer to achieve bit rates of up to 6 Mbit/s. This shows that the concept of a stream-oriented light-weight AAL in addition to provide better network QoS is easy to implement and does not require many resources. This situation is allowed by the simple mechanisms proposed for the MAAL. Taking into account, from the beginning of the design, the specifics of interactive multimedia applications allows to derive simple but efficient mechanisms.

The transmission of data files, according to the ATM Forum VoD specification [155], containing CBR MPEG-2 encoded video streams over the MAAL prototype has been performed successfully. The results show improvements when FEC is used as already shown via simulation in Chap. 5. The cell loss processes sampled within the ATM LAN also agree with the ones obtained in Chap. 5 and explain the similar CLR results achieved in both simulated and real environments. The PSD figures also confirm the validity of the low traffic source assumption approach taken in Chap. 4 to design the MAAL.

The performance and QoS results achieved by the MAAL prototype show that to improve the transmission of interactive multimedia streams it is not necessary to implement complex protocol layers such as AAL1. Simple mechanisms that take into account multimedia application requirements and in particular user requirements offer several benefits at the network and user level without sacrificing bandwidth or protocol complexity.

172

Chapter 8

Final Conclusions

8.1 Achievements

The traditional boundaries between the telecommunications, computer and entertainment industries are blurring. However, what seems to be convergence in reality is not. Each of the industries approach audiovisual communications from different technological perspectives, with each industry providing its own, often incompatible standards and solutions for similar applications. This lack of integration has been addressed in this work.

One of the contributions of this thesis is the development of a set of multimedia protocol layers designed with integration as a target. The layers had to efficiently interface multimedia applications to ATM networks. Thus, the user and the multimedia applications requirements are the criteria used for the design of the protocol layers. Unlike traditional data communications, in multimedia the user is directly involved in the outcome of the service. The user judges whether the service is according to his expectations or demands. Therefore, one of the key design principles of the protocol layers has been to take into account user perception. This imposes the protocol layers to be able to handle VBR traffic. To guarantee a near constant quality video, the encoders must deliver VBR data streams. The utilization of constant data rate sources, albeit easier to handle from a networking point of view, requires from the encoder to modify the quality of the compressed video data to fulfill the rate constraint.

Two layers have been proposed:

- a generic multimedia ATM adaptation layer, called MAAL, which provides *efficient low latency and reliable transmission* for both CBR and VBR data streams

- a codec specific network adaptation layer or NAL that fully exploits the functions of the MAAL. It provides a selective error correction based on the *perceptual relevance* of the data.

Providing a perceptual error protection mechanism requires an a priori knowledge of the data to transmit. This entails a dependency on the type of codec. Therefore the rationale to develop two layers has been the need to split functionality into generic network functions provided by the MAAL and codec specific functions provided by the NAL.

The timing constraints being a key issue in interactive multimedia communications, both layers provide simple low delay functions. The proof-of-concept prototype software implementation of the MAAL proves that the complexity cost is not an issue. The set of functions developed for both layers are minimal because they target a single class of time-constrained applications. Thus the layers are light and consequently easy to implement.

From a networking perspective, the efficiency of the proposed layers is explained by the loss profiles observed in the experiments. We have shown for a new class of sources, VBR, that the *loss process* of a single stream using a relatively small fraction of the link rate and multiplexed with background traffic coming from a large number of sources, can be approximated by a *uniformly distributed process.* The consequences of this observation are extremely important. Under this loss profile, FEC achieves the best recovery performance. The protection efficiency is not directly related to powerful and complex FEC codes but to sparse losses easy to correct with minimal redundancy. Simultaneously, uniformly distributed loss processes are the worst case for packet-oriented mechanisms such as AAL5's packet discard.

This thesis has also contributed towards the development and understanding of user-oriented metrics for multimedia communications. The innovative use of the recently developed *perceptual quality metric* called Moving Pictures Quality Metric or MPQM for *network performance evaluation* is a first step towards the still missing *user-oriented* metrics for multimedia.

The usage of phsychophysics for network performance assessment has brought to the fore how diverging the provider and user-oriented metrics could be. The experimental results showed that in average unexpectedly high CLRs are tolerated by the user. Indeed, the transmission of video streams over the NAL-MAAL layers and damaged by CLR values of up to 10^{-3}, generally considered as unacceptable for ATM connections, showed that the average perceptual quality is very close to the original encoding quality. This result could not have been derived without perceptual metrics.

The implications of this observation are paramount. The high loss tolerance of the perceptually protected video streams combined to the usage of VBR streams allows to improve the network utilization by means of statistical multiplexing. Economically, this may be interesting for both the network provider and the user because it allows to *reduce the cost per communication* by accepting more connections without reducing the perceived quality which in the end is one of the main goals of research in telecommunications.

In fact, only the combined usage of MPQM and network QoS metrics has been able to *unveil the economic potential* of perceptually and selectively protected streams.

8.2 Extensions

The extensions of this work are multiple.

The development of a codec specific NAL is close to the concept of application level framing that advocates a high level of integration between the application and the protocol layers in order to optimize network services. The difference between our concept and ALF is that although NALs are codec specific, they still provide functions generic enough to be used by a large number of applications. The growing number of multimedia applications using the same video codecs proves it.

Also, the NAL concept can be pushed further to offer a complete protection mechanism totally independent of the underlying network technology. Moving the FEC data generation into the NAL, would result in a self-contained layer offering the PSIP algorithm and the FEC code. The advantage would be its applicability to any type of network, being ATM, IP based or other. The drawback would be a loss in recovery efficiency and a higher overhead.

The JAVA language opens new possible extensions to the concept of network adaptation. Given a certain degree of specificity it is possible to develop NALs as downloadable software protocol layers or "proclets". These proclets could be implemented in the sender only and downloaded into the receiver. After usage, the NAL could be erased from the receiver as it is envisaged in MPEG-4. This scheme has the advantage of avoiding the cluttering of the protocol layer stack with several NALs, one for each existing or new audiovisual compression algorithm.

The promising results shown by the proposed layers let foresee another domain of application: the transmission over wireless networks. One of the characteristics of wireless networks is the relatively low quality of the channels. High error rates generally due to fading effects are common. However, the loss profiles exhibit a much burstier behavior. The proposed protocol layers have not been designed with such loss processes in mind. Therefore, further research would be necessary to accomodate such loss profiles. Still, the flexibility offered by both the MAAL and the NAL make them good candidates to adapt to the conditions offered by wireless communications.

Using Psychophysics as a measurement tool for network performance paves the way towards more in depth studies of the techniques for mapping user-to-network QoS parameters. The current work has considered only the evolution of the perceptual quality as a function of CLR. More advanced studies should also take into account the encoding bit rate as a parameter. The result of such models will be a three-dimensional curve describing the influence of each of these major parameters. The evolution of the perceptual quality in this three-dimensional space will be an essential tool to fine tune network QoS parameters according to user expectation.

In the end, the extensions described here, and in particular the proclets and the 3D curves, may be considered as tools. These tools will become more and more important to network operators because they will allow to improve quality and network utilization. The added-value that such tools will provide may be the key to offer better services at lower prices. Therefore, mastering these tools could become an economical asset for providers and network operators.

Ultimately, the proposed framework could lead to the development of software libraries, in particular for the NALs that could be included with hardware codecs or be sold as separate quality enhancement products for interactive multimedia communications.

Appendix A

Multimedia AAL Specification

We derive in this appendix a specification of the functions and data structure of the MAAL. It is not our intention to develop a full ITU-T alike AAL recommendation but rather to describe in a formal way the most important aspects that make this AAL suitable for real-time interactive multimedia applications. This section gives also *guidelines* for a possible AAL implementation.

A.1 Multimedia AAL

A.1.1 Services Provided by the Multimedia AAL

The layer services provided by the MAAL to the AAL user are:

- transfer of constant size service data units with a variable source bit rate
- indication of lost or errored information which is not recovered by the AAL.

A.1.2 Primitives Between MAAL and the AAL User

A.1.2.1 General

At the AAL-SAP, the following primitives will be used between the MAAL and the AAL user:

- from an AAL user to the AAL, AAL-UNITDATA-REQUEST
- from the AAL to the AAL user, AAL-UNITDATA-INDICATION.

An AAL-UNITDATA-REQUEST primitive at the local AAL-SAP results in an AAL-UNITDATA-INDICATION primitive at its peer AAL-SAP.

A.1.3 Definition of Primitives

The results of the precedent sections demonstrate that the utilization of a stream-oriented AAL provides much better results in terms of both network QoS and user perceived quality. This section specifies the functions and data structure of the MAAL.

A.1.3.1 AAL-UNITDATA-REQUEST

AAL-UNITDATA-REQUEST (DATA[mandatory], FEC-REQUEST[optional])

The AAL-UNITDATA-REQUEST primitive requests the transfer of the AAL-SDU, i.e contents of the DATA parameter, from the local AAL entity to its peer entity. The length of the AAL-SDU is constant and the time interval between two consecutive primitives should be variable. The length parameter is fixed by the AAL user at connection setup.

A.1.3.2 AAL-UNITDATA-INDICATION

AAL-UNITDATA-INDICATION (DATA[mandatory], STATUS[mandatory])

An AAL user is notified by the AAL that the AAL-SDU, i.e. contents of the DATA parameter, from its peer are available. The length of the AAL-SDU should be less or equal to the length negotiated at connection setup depending on the utilization or not of dummy cell insertion. The time interval between two consecutive primitives should be variable.

A.1.4 Definition of Parameters

A.1.4.1 FEC-REQUEST Parameter

The FEC-REQUEST parameter is used to notify the AAL the number of FEC cells to be generated to protect the AAL-SDU. The FEC-REQUEST parameter values are bounded by the maximum number of FEC cells allowed by the FNI field of the AAL SAR-PDU. As an example if the FNI field is two bits long, the maximum value of the FEC-REQUEST parameter is four. The default value of FEC-REQUEST is zero meaning that no redundancy cells must be generated for the AAL-SDU. If the upper layer is unable to pass the FEC-REQUEST parameter it will be fixed at connection setup. If no value is given, then the default value is used.

A.1.4.2 STATUS Parameter

The STATUS parameter identifies that the DATA is judged to be errored or not after error recovery if applicable. The STATUS parameter has two values:

VALID

INVALID

The INVALID status could also imply that part of the DATA is a dummy value. The choice of dummy value depend on the type of AAL service provided and is agreed at connection setup.

A.2 Segmentation and Reassembly (SAR) Sublayer

A.2.1 Functions of the SAR Sublayer

The SAR sublayer functions are performed on an ATM-SDU basis.

1. Mapping between CS-PDU and SAR-PDU: The SAR sublayer at the transmitting end accepts 47 octet block of data from the CS, and prepends a one octet SAR-PDU header to each block to form the SAR-PDU as depicted in Fig. A.1.

The SAR sublayer at the receiving end receives 48 octet block of data from the ATM layer, and then separates the SAR-PDU header. The 47 octet block of SAR-PDU payload is passed to the CS.

2. Sequence numbering: Associated with each SAR-PDU payload, the SAR sublayer receives a sequence number value from the CS. At the receiving end, it passes the sequence number value to the CS. The CS uses these sequence number values to detect lost or misinserted SAR-PDU payloads.

3. Error protection: The SAR sublayer protects the SAR-PDUs against cell losses. It informs the receiving CS when missing SAR-PDUs are detected.

A.2.2 SAR Protocol

The SAR-PDU header together with the 47 octets payload form the SAR-PDU. The SAR-PDU header comprises two fields: the sequence number field and the FEC Number Indication (FNI) field. The size and positions of the fields in the SAR-PDU are given in Fig. A.1

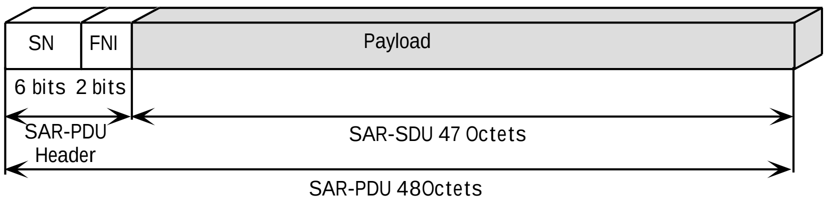


Figure A.1: *MAAL SAR-PDU structure.*

A.2.2.1 Sequence Number (SN) field

Our studies have shown that for any combination of correlated and uncorrelated foreground and background traffic, the number of consecutive lost cells is relatively small with high percentages of single cell losses. We have shown that a large percentage of losses are single cell losses. Also the number of cell losses per packet is generally small. We therefore use a sequence number field of 6 bits which allows to cover up to 64 consecutive cell losses.

1. Operations at transmitting end:

   the transmitter computes the sequence number on a modulo 64 basis. After calculation the transmitter inserts the value in the SN field.

2. Operations at receiving end:

   the receiver operates in detection mode. It monitors each SAR-PDU header by checking the SN field continuity. If a header error is detected the receiver conveys the sequence number count to the CS together with SN check status (valid or invalid).

A.2.2.2 FEC Number Indication (FNI) field

The results of the precedent sections show that at the cell level using small packet sizes reduces the overhead and requires a relative small number of FEC cells to achieve a high degree of reliability. We use a FNI field of 2 octets able to signal the existance of a maximum of 4 FEC cells.

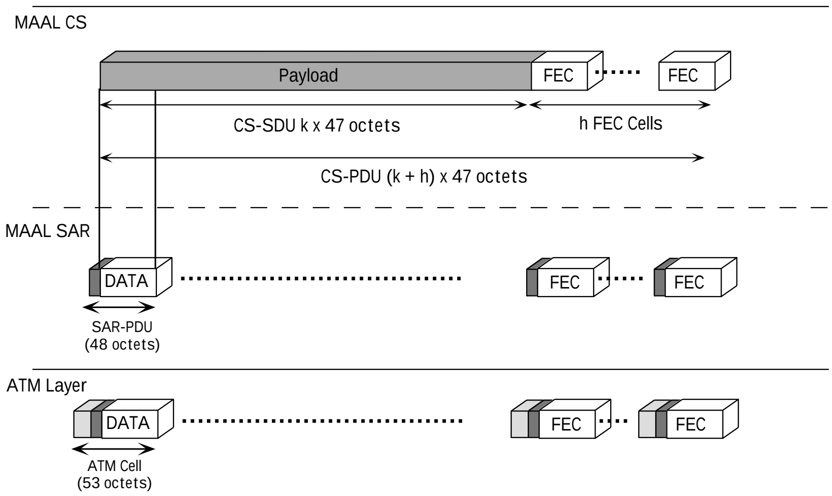


Figure A.2: *MAAL CS-PDU structure.*

1. Operations at transmitting end:

   The transmitter inserts into the FNI field the number of FEC cells to be generated for the corresponding CS-SDU according to the FNI-REQUEST parameter of the AAL-UNITDATA-REQUEST primitive as described in Sec. A.1.3.1. The number of FEC cells is generated on a CS-SDU basis and may be different for each CS-SDU. The CS-SDU with the FEC cells constitutes the CS-PDU as shown in Fig. A.2.

2. Operations at receiving end:

   The receiver monitors the FNI field of the cells pertaining to a CS-PDU. The FNI information fixes the number of cells of a CS-PDU. The FNI informationis used for CS-PDU delineation. The CS-PDU is passed to the CS where cell loss recovery may be performed on a SAR-SDU basis.

A.2.3 Convergence Sublayer (CS)

The CS may include the following functions:

- processing of sequence count is performed at this sublayer. The sequence count value and its error check status provided by the SAR sublayer are used by the CS to detect loss and misinsertion. Further handling of lost and misinserted cells is also performed in this sublayer according to FNI information
- generation of FEC cells is performed according to the FNI-REQUEST parameter which indicates the number of FEC cells to be appended to the CS-SDU
- handling of loss and misinserted cells is performed at this sublayer. The sequence values are processed to detect lost and misinserted cells. Misinserted cells are discarded.

For some AAL users it is required to maintain bit count integrity of the AAL user information. In this case lost cells detected are replaced by inserting the appropriate number of dummy SAR-PDU payloads.

According to FNI information, cell losses may be corrected. For some AAL users the corrected cells with the original information replace the inserted dummy payloads. When correction is not possible, then the errored CS-PDU with or without the dummy information is passed to the upper layer.

A.2.4 Correction Method for Cell Losses

The correction method uses forward error correction. FEC uses Reed-Solomon erasure codes (RSE). The number of FEC cells equals the number of recoverable cell losses.

In the transmitting CS, the FEC cells are appended to the CS-SDU to form a CS PDU. The number of appended FEC cells is indicated in the FNI of each transmitted cell of the corresponding CS-PDU. The CS-PDU is forwarded to the SAR sublayer. The SAR-PDU payloads contain the information required to recover the SAR-PDU payloads of the user information. The number of redundancy cells generated is variable for each CS-SDU. It is user dependent and is fixed by the FEC-REQUEST parameter of the AAL-UNITDATA-REQUEST primitive as described in Sec. A.1.3.1.

The loss of one SAR-PDU payload in the CS-PDU implies one erasure. Erasures may correspond to dummy cell payloads in the CS-PDU when requested by the AAL user.

A.2.5 Error Correction Delay

The FEC cells will be generated by RSE mechanism no interleaving will be applied therefore achieving a maximum average delay of:

$$Delay_{RSE} = \frac{k}{Cell\ rate} sec, \tag{A.1}$$

where k is the number of FEC cells per packet.

Appendix B

MPEG-2 Network Adaptation Layer Specification

The simulation results of Sec. 6.7.2 show that the transmission of MPEG-2 video with a PSIP mechanism achieves much better performance from the user perspective. A PSIP mechanism requires a NAL able to perform the selection of data to be protected. The choice of transmitting constant size AAL-SDUs requires from the NAL to perform a segmentation of the NAL-SDUs prior to apply the PSIP algorithm. The remainder of this chapter specifies the interface of an MPEG-2 specific NAL with the MAAL described in Chap. 5. This specification may be used as a guideline for a real implementation.

B.1 MPEG-2 Specific Network Adaptation Layer

B.1.1 Services Provided by the MPEG-2 NAL

The layer services provided by the NAL to the NAL user are:

- transfer of variable size service data units with a variable source bit rate
- generation of FEC request information
- segmentation of application data packets into constant length NAL protocol data units
- alignement of data packets onto MAAL service data units.

B.1.2 Primitives Between NAL and the NAL User

At the NAL-SAP, the following primitives will be used between the NAL and the NAL user:

- from a NAL user to the NAL, NAL-UNITDATA-REQUEST
- from the NAL to the NAL user NAL-UNITDATA-INDICATION.

An NAL-UNITDATA-REQUEST primitive at the local NAL-SAP results in an NAL-UNITDATA-INDICATION primitive at its peer NAL-SAP.

B.1.3 Definition of Primitives

B.1.3.1 NAL-UNITDATA-REQUEST

NAL-UNITDATA-REQUEST(DATA[mandatory])
The NAL-UNITDATA-REQUEST primitive requests the transfer of the NAL-SDU, i.e. the contents of the DATA parameter, from the local NAL entity to its peer entity. The length of the NAL-SDU is variable and the time interval between two consecutive primitives should be variable.

B.1.3.2 NAL-UNITDATA-INDICATION

NAL-UNITDATA-INDICATION(DATA[mandatory], STATUS[optional])
A NAL user is notified by the NAL that the NAL-SDU, i.e. contents of the DATA parameter, from its peer is available.

B.1.4 Definition of STATUS Parameter

The STATUS parameter identifies that the DATA is judged to be errored or not. The STATUS parameter has two values:

VALID

INVALID

The INVALID status could also imply that part of the data is a dummy value. The choice of dummy value depend on the type of service provided by the AAL.

B.2 Functions of the Network Adaptation Layer

The NAL functions are performed on a AAL-SDU basis.

1. Mapping between NAL-SDU and AAL-SDU: The NAL at the transmitting end accepts data packets of fixed size from the NAL user and performs a padding function when the size of the user data packet does not match the AAL-SDU size negotiated at connection setup[1].

   The NAL at the receiving end receives one NAL-PDU of fixed length from the AAL. If byte stuffing is performed it is not signaled to the NAL user. The NAL-SDU is passed as is to the NAL user.

2. Error protection: The NAL identifies MPEG-2 syntactic information as described in [16] and asks for FEC protection via the FEC-REQUEST parameter according to the PSIP algorithm agreed at connection setup. Section 6.5.2.1 gives an example of a PSIP algorithm for MPEG-2.

B.2.1 Segmentation and Reassembly

The MPEG-2 system Transport Stream system layer delivers constant length TS packets at its interface. Thus, the MPEG-2 NAL does not need to perform a packet segmentation. This is not the case with all the available encoders. An MJPEG encoder delivers variable size packets composed of a single frame. In that case, the MJPEG specific NAL will perform

[1]This may be required at the end of a sequence when the number of TS packets does not match the negotiated AAL-SDU length

a packet segmentation to obtain constant length NAL-PDUs. This also requires from the NAL to perform a delineation function to inform the receiver end when a complete packet has been transmitted which leads to some overhead.

The only case when MPEG-2 requires byte stuffing is at the end of a sequence. Consider as an example that the agreed size is 376 octets AAL-SDUs. This means two TS packets per NAL-PDU. If the total number of TS packets that constitute the sequence is odd then the last NAL-PDU will be 188 ocets long. The NAL performs a byte stuffing by adding 188 dummy octets to align the NAL-PDU to the AAL-SDU size. It is not required to provide a specific field for this function since the *dummy TS packet* will be discarded by the system layer. The overhead generated by this stuffing mechanism is bounded by:

$$overhead = \frac{AAL - SDU\ length - 188\ octets}{Sequence\ length}. \tag{B.1}$$

Since neither delineation nor alignment functions are required the NAL does not add any header or trailer to the NAL-SDU.

B.2.2 Error Protection

The main function of the NAL is the error protection. The NAL does not generate the error protection information itself. The NAL generates FEC-REQUEST signals passed to the AAL with the AAL-SDU in order to selectively protect NAL-PDUs according to an algorithm that could be predefined or loaded at connection setup.

- Operations at transmitting end:

  The transmitter analyzes the TS packets of a NAL-SDU to identify syntactic information as specified in [16]. According to a priority table, a FEC-REQUEST containg the number of FEC cells to append is generated. When more than one header of different priorities are identified, the highest priority is used.

- Operations at receiving end:

  The receiver NAL is transparent in this specific case[2]. The FEC-REQUESTS are handled by the receiver's MAAL. When byte stuffing is performed, the NAL-SDUs are passed without notification to NAL user.

[2]This will not be the case when the NAL performs user data segmentation and reassembly.

Appendix C

MAAL Software Implementation Structures

C.1 MAAL Structures description

This appendix contains the full description of the structures defined in the MAAL classes. These structures are used internally as well as for the export of values to the user.

C.1.1 Structures of the Common Class

```
typedef enum {
     first_cell,
     data,
     complete
} recv_state
```

This structure is part of the decode_PDU internal function. It summarizes the different receiver states which are the following:

- first_cell: the function waits for the first cell of a group (PDU). It ignores any other cell
- data: waits for a data cell
- complete: the PDU is complete and can be send to the MAAL user.

```
typedef struct {
int                PDU_size;
int                mtu;
Atm_endpoint       endpoint;
Atm_qos_sel        qos_local;
Atm_qos            qos_destination;
bool               FEC_enabled;
} ATM_params;
```

This structure is used to pass the connection parameters to the user. Not all internal parameters are included in the structure because they are not necessary. The following are proposed:

- PDU_size: the size of the PDU used in the ongoing connection. It is given in multiples of 47 octets
- mtu: Maximum Transmission Unit. The maximum transmission size unit given in octets
- endpoint: gives the access point number of the remote site for the ongoing connection
- asap: the local application service access point
- qos_local: the local quality of service
- qos_destination: the remote quality of service
- FEC_enabled: true if the FEC is being used in the ongoing connection.

```
typedef struct {
int                fd;
int                PDU_size;
int                mtu;
Atm_endpoint       endpoint;
Atm_qos_sel        qos_local;
Atm_qos            qos_destination;
char               current_seq_nb;
bool               FEC_enabled;
int                FEC_cells;
char               *FEC_buffer;
char               *recvBuf;
char               *sendBuf;
char               total_FEC;
char               *PDU_fragment;
bool               *PDU_fragment_exists;
bool               *PDU_fragment_is_first;
bool               *PDU_cells_state;
int                write_to_receptor;
int                read_frm_receptor;
bool               use_receptor;
} ATM_connection;
```

This structure contains the full set of parameters used to describe and ATM connection. The parameters are described next:

- fd: file descriptor for the current connection. This parameter is given by atm_open
- PDU_size: the size of the PDU used in the ongoing connection. It is given in multiples of 47 octets
- mtu: Maximum Transmission Unit. The maximum transmission size unit given in octets
- endpoint: gives the access point number of the remote site for the ongoing connection
- asap: the local application service access point
- qos_local: the local quality of service
- qos_destination: the remote quality of service

- current_seq_nb: the actual sequence number of the last received cell
- FEC_enabled: true if the FEC is enabled for the ongoing connection
- FEC_cells: number of FEC cells for each PDU
- FEC_buffer: the buffer allocated for the incoming FEC cells. The buffer is allocated at the beginning of the connection to avoid a new allocation every PDU arrival
- recvBuf: buffer of size mtu used to store the cells received via the FORE routine atm_recv_null. It is allocated at the beginning of the connection
- sendBuf: buffer of size mtu used to store the cells to be transmitted via the FORE routine atm_send_null. It is allocated at the beginning of the connection
- total_FEC: the total amount of FEC cells for a given PDU. It is equal to 0 if FEC is disabled
- PDU_fragment: a PDU fragment from the current received PDU
- PDU_fragment_exists: is true if a fragment of a PDU exists. In that case it is used to construct the current PDU
- PDU_fragment_is_first: is true if the received cell is the first of a PDU
- PDU_cells_state: state of the received cells. It indicates which cells have been successfully received. This attribute is used to perform cell recovery
- write_to_receptor: the pipe descriptor to write to the child process the ATM cells
- read_frm_receptor: the pipe descriptor to read from the child process the ATM cells
- use_receptor: is true if the server uses a child process to received the cells.

C.1.2 Constants used in the MAAL Classes

The constants used in the three classes are described in this section. A modification of these constants entails a modification of the behavior of the MAAL without any modifcation of the functions.

- AAL0_BYTES: indicates the number of bytes of an AAL0-SDU equal to a MAAL-PDU. The value is set to 48
- MAAL_BYTES: indicates the number of bytes of an MAAL cell payload. The value is set to 47
- MAX_SEQ_NUM: defines the maximum sequence number alowed in a cell. The value is set to 32 (5 bits)
- SEQ_NUM_MASK: mask allowing to extract the sequence number from the MAAL header. The value is set to 0x1f hexadecimal
- TYPE_MASK: mask allowing to extract the cell type (DATA, FEC). The value is set to 0x60
- FIRST_CELL_MASK: mask allowing to extract the first cell tag of the PDU. The value is set to 0x80
- FEC_CELL: defined the FEC cell type (bits 5 and 6 = 1). The value is set to 96

- DATA_CELL: defines the DATA cell type (bits 5 and 6 = 0). The value is set to 0
- ERROR_CELL: defines the unknown cell type (bit 5 = 0 and bit 6 = 1). The value is set to 64
- FIRST_CELL: defines the first cell tag. The value is set to 128
- TIMEOUT_SEC: defines the maximum time in seconds the recv_PDU function waits before returning a MAAL_ERROR_TIMEOUT. The value is set to 1
- TIMEOUT_USEC: defines the maximum time in μs the recv_PDU function waits before returning a MAAL_ERROR_TIMEOUT. The value is set to 0 (disabled)
- END_CONNECTION: defines the message send by the parent process to kill the child process or release message. The value is set to "END"
- END_SIZE: defines the size in bytes of the connection release message. The value is set to 4
- BATCH_SIZE: defines the size of the buffer allocated for the atm_recv_null function when using a child process. The value is set to 5.

Bibliography

[1] ITU-T, editor, *Recommendation I.320 ISDN Protocol Reference Model*, ITU-T, Geneva, Nov. 1993.

[2] Lixia Zhang, Stephen Deering, Deborah Estrin, Scott Shenker and Daniel Zappala, "RSVP: A New Resource ReSerVation Protocol", *IEEE Network*, vol. , pp. 8–18, Sep. 1993.

[3] Scott O. Bradner and Allison Mankin, *IPng Internet Protocol Next Generation*, Addison-Wesley, 1995.

[4] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", Technical Report RFC 1889, Internet Engineering Task Force, 1996.

[5] H. Leopold, G. Blair, A. Campbell, G. Coulson, P. Dark, F. Garcia, D. Hutchinson, N. Singer, and N. Williams, "Distributed Multimedia Communication System Requirements", OSI 95 Consortium, May 1992.

[6] Ralf Steinmetz and Klara Nahrstedt, *Multimedia: Computing, Communications & Applications*, Prentice-Hall, 1995.

[7] M. Goldstein, N.Björkmann, and A. Latour, "Cell Loss Effects on Perceived Audio/Video Quality and Quality of Service (QOS)", *in 6th Workshop in Packet Video*, Portland, USA, Oct. 1994.

[8] Ralf Steinmetz, "Data Compression in Multimedia Computing - Principles and Techniques", *Multimedia Systems*, vol. , 1994.

[9] Ralf Steinmetz, "Data Compression in Multimedia Computing - Standards and Systems", *Multimedia Systems*, vol. , 1994.

[10] Joan L. Mitchell, William B. Pennebaker, Chadd E. Fogg, and Didier J. LeGall, *MPEG Video Compression Standard*, Digital Multimedia Standards. Chapman & Hall and International Thomson Publishing, 1996.

[11] ITU-T, editor, *Recommendation H.261 Video Codec for Audiovisual Services at p x 64 kbits*, ITU-T, Geneva, 1993.

[12] ITU-T, *Recommendation T.81*, ITU-T, Geneva, Mar. 1993.

[13] ISO/IEC JTC-1, editor, *ISO/IEC 11172-2, Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - Part 2: Video*, ISO/IEC JTC-1, 1993.

[14] ISO/IEC JTC-1, editor, *ISO/IEC 11172-1, Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - Part 1: Systems*, ISO/IEC JTC-1, 1993.

[15] ISO/IEC JTC-1, editor, *ISO/IEC 11172-3, Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - Part 3: Audio*, ISO/IEC JTC-1, 1993.

[16] ISO/IEC JTC-1, editor, *ISO/IEC 13818-1, Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1: Systems Specification*, ISO/IEC JTC-1, 1994.

[17] ISO/IEC JTC-1, editor, *ISO/IEC 13818-2, Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video*, ISO/IEC JTC-1, 1994.

[18] ISO/IEC JTC-1, editor, *ISO/IEC 13818-3, Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 3: Audio*, ISO/IEC JTC-1, 1994.

[19] ITU-T, editor, *Recommendation H.262 Coding of Moving Pictures*, ITU-T, Geneva, July 1995.

[20] ISO/IEC JTC-1, editor, *ISO/IEC 13818-1, Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1: Systems Specification*, In JTC-1 [16], 1994.

[21] ITU-T, editor, *Recommendation H.263 Video Coding for Low Bit Rate Communications*, ITU-T, Geneva, Mar. 1996.

[22] ISO-IEC/JTC1SC29/WG11, "Requirements for the MPEG-4 syntactic description language", Technical report, ISO, 1995.

[23] ISO-IEC/JTC1SC29/WG11, "MPEG-4 proposal package description (ppd)", Technical report, ISO, 1995.

[24] M. Garret and W. Willinger, "Analysis Modeling and Generation of Self-Similar VBR Video Traffic", *in ACM SIGCOMM 94*, London, 1994. ACM, Academic Press.

[25] Oliver Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems", Technical report, COST 242, 1995.

[26] W. Willinger W. Leland, M. Taqqu and D. Wilson, "On the self-similar nature of ethernet traffic", *IEEE/ACM Transactions on Networking*, pp. 1–15, February 1994, Extended Version.

[27] Jean-Yves Le Boudec, "The asynchronous transfer mode: a tutorial", *Computer Networks and ISDN Systems*, vol. 24, pp. 279–309, 1992.

[28] Martin de Prycker, *Asynchronous Transfer Mode: Solution for broadband ISDN*, Ellis Horwood, third edition, 1995.

[29] ITU-T, editor, *Recommendation I.432 B-ISDN User-Network Interface - Physical Layer Specification*, ITU-T, Geneva, Mar. 1993.

[30] T1X1.2 Working Group on Digital Transmission Network Architecture, "A Technical Report on a Comparison of SONET (Synchronous Optical NETwork) and SDH (Synchronous Digital Hierarchy)", Technical report, ANSI, 1993.

[31] ITU-T, editor, *Recommendation I.361 B-ISDN ATM Layer Specification*, ITU-T, Geneva, Mar. 1993.

[32] ITU-T, editor, *B-ISDN Asynchronous Transfer Mode Functional Characteristics*, ITU-T, Geneva, Mar. 1993.

[33] ITU-T, editor, *Recommendation I.363 B-ISDN ATM Adaptation Layer (AAL) Specification*, ITU-T, Geneva, Mar. 1993.

[34] ITU-T Study Group 13, editor, *Recommendation I.363.X B-ISDN ATM Adaptation Layer Specification, Types 1 and 2*, ITU-T, Geneva, Mar. 1994, Draft.

[35] ITU-T Study Group 13 Question 6, "Report of the q.6 meeting", May 1996.

[36] ITU-T Study Group 13, editor, *Recommendation I.362 B-ISDN ATM Adaptation Layer Functional Description*, ITU-T, Geneva, May 1996, Frozen.

[37] ITU-T Study Group 13, editor, *Recommendation I.363.1 B-ISDN ATM Adaptation Layer Specification, Types 1 and 2*, ITU-T, Geneva, July 1995, Draft.

[38] Katsuyuki Yamazaki and Yoshikazu Ikeda, "Connectionless cell switching schemes for broadband ISDN", *in XIII International Switching Symposium*, vol. VI, pp. 5–10, Stockholm, Sweden, June 1990.

[39] Jean-Yves Le Boudec, Daniel A. Pitt, and Hong Ling Truong, "ATM connectionless service using computer networking methods", *in International Conference on Computer Communications ICCC'92*, Sep. 1992.

[40] ITU-T, editor, *B-ISDN ATM Layer Cell Transfer Performance*, ITU-T, Geneva, Nov. 1993.

[41] ITU-T, editor, *Recommendation I.371 Traffic Control and Congestion Control in B-ISDN*, ITU-T, Geneva, July 1995, Draft.

[42] The ATM Forum Technical Committee, *Traffic Management Specification Version 4.0*, The ATM Forum Traffic Management Working Group, Apr. 1996.

[43] The ATM Forum, "AMS Phase 2 Living List", Oct. 1996.

[44] Thomas Stock and Reto Gruenenfelder, "Frame Loss vs. Cell Loss in ATM Concentrators and Multiplexing Units", *in EFOC&N'94*, 1994.

[45] Israel Cidon, Roch Guerin, Asad Khamisy, and Kumar N. Sivarajan, "Cell versus Message Level Performances in ATM Networks", *in First International ATM Traffic Expert Symposium*, Basel, Switzerland, Apr. 1995.

[46] J. M. Simmons and R. G. Gallager, "Design of Error Detection Schemes for Class C Service in ATM", *IEEE/ACM Transactions on Networking*, vol. 2, Feb. 1994.

[47] J.-P. Leduc, *Digital Moving Pictures - Coding and Transmission on ATM Networks*, vol. 3 of *Advances in Image Communications*, Elsevier, 1994.

[48] Jean Walrand and Pravin Varaiya, *High-Performance Communication Networks*, Morgan Kaufmann, 1996.

[49] George Kesidis, Jean Walrand, and Cheng-Shang Chang, "Effective bandwidths for multiclass markov fluids and other atm sources", *IEEE/ACM Transactions on Networking*, vol. 1, pp. 424–428, 1993.

[50] Roch Guerin, Hamid Ahmadi, and Mahmoud Naghshineh, "Equivalent bandwidth and its application to bandwidth allocation in high speed networks", *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 968–981, 1991.

[51] W. Verbiest, L. Pinoo, and B. Voeten, "Statistical multiplexing of variabke bit rate video sources in asynchroonous transfer mode networks", *in ICC'88*, 1988.

[52] Andrea Baiocchi, Nicola Blefari Melazzi, Narco Listanti, Aldo Roveri, and Roberto Winkler, "Loss performance analysis of an ATM multiplexer loaded with high-speed on-off sources", *IEEE Journal on Selected Areas in COmmunications*, vol. 9, Apr. 1991.

[53] Phuoc Tran-Gia and Hamid Ahmadi, "Analysis of a discrete-time $G^{[x]}/D/1--S$ queueing system with applications in packet-switching systems", *in IEEE INFOCOM'88*, Mar. 1988.

[54] San-Qi Li, "A general solution technique for discrete queueing analysis of multimedia traffic on ATM", *IEEE Transactions on Communications*, vol. 39, pp. 1115–1132, 1991.

[55] B. Maglaris, Dimitris Anastassiou, Prodip Sen, Gunnar Karlsson, and John D. Robbins, "Performance models of statistical multiplexing in packet video communications", *IEEE Transactions on Communications*, vol. 36, pp. 834–844, 1988.

[56] Mark W. Garret, *Contributions Toward Real-Time Services on Packet Switched Networks*, PhD thesis, Columbia University, 1993.

[57] Henning Schulzrinne, James F. Kurose, and Donald Towsley, "Loss correlation for queues with bursty input streams", *in International Conference on Communications ICC'92*, vol. 1, pp. 219–224, Chicago, Illinois, June 1992. IEEE.

[58] I. Norros and J. T. Virtamo, "Who loses cells in the case of burst scale congestion?", in A. Jensen and V.B. Iversen, editors, *ITC-13*. Elsevier Science Publishers, 1991.

[59] David M. Cohen, Daniel P. Heyman, and Amy R. Reibman, "Comparisons among Models of Cell Losses for Video Conferences", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[60] Reto Grünenfelder, Cesar Gallego, and Yves Gachoud, "Measurememnt and Characterization of Cell Loss in ATM Switches", *in GLOBECOM'94*, pp. 1075–1079, 1994.

[61] David M. Cohen and Daniel P. Heyman, "A Simulation Study of Video Teleconferencing Traffic in ATM Networks", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, 1993.

[62] Xavier Garcia Adanez and Thomas Stock, "Interworking concepts for multimedia services in Broadband-ISDN – first report", Technical Report KWF 2760.1, EPFL, Dec. 1994.

[63] Andre B. Bondi and Wai-Sum Lai, "The influence of cell loss patterns and overheads on retransmission choices in broadband isdn", *Computer Networks and ISDN Systems*, vol. 26, Jan. 1994.

[64] Christian J. van den Branden Lambrecht and Olivier Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", *in Proceedings of the SPIE*, vol. 2668, pp. 450–461, San Jose, CA, January 28 - February 2 1996, available on `http://tcomwww.epfl.ch/~verscheu`.

[65] Ismail Dalgic and Fouad A. Tobagi, "Glitches as a Measure of Video Quality Degradation Caused by Packet Loss", *in 7th Workshop in Packet Video*, Brisbane, Australia, Mar. 1996.

[66] Arthur A. Webster, Coleen T. Jones, Margaret H. Pinson, Stephen D. Voran, and Stephen Wolf, "An objective video quality assessment system based on human perception", *in SPIE Human Vision, Visual Processing, and Digital Display IV*, vol. 1913, pp. 15–26, San Jose, CA, Feb. 1993.

[67] Serge Comes, *Les traitements perceptifs d'images numerisées*, PhD thesis, Université Catholique de Louvain, 1995.

[68] Christian J. van den Branden Lambrecht, "A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications", *in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 2293–2296, Atlanta, GA, May 7-10 1996, available on `http://ltswww.epfl.ch/~vdb`.

[69] M. Ardito, M. Barbero, M. Stroppiana, and M. Visca, "Compression and Quality", *in Proceedings of the International Workshop on HDTV 94*, Brisbane, Australia, October 1994.

[70] Pär Lindh and Christian J. van den Branden Lambrecht, "Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences", *in Proceedings of the International Conference on Image Processing*, Lausanne, Switzerland, September 16-19 1996, available on `http://ltswww.epfl.ch/~vdb`.

[71] David J. Heeger and Patrick C. Teo, "A Model of Perceptual Image Fidelity", *in Proceedings of the International Conference on Image Processing*, pp. 343–345, Washington, DC, October 23-26 1995.

[72] Mohammad Ghanbari, "An Adapted H.261 Two-Layer Video Codec for ATM Networks", *IEEE Transactions on Communications*, vol. 40, Sep. 1992.

[73] M. Khademi and F.J. Paoloni, "Layered Video Communications in an ATM Environment", *in ICCS'94*, Singapore, 1994.

[74] R. E. Van Dyck, C. H. Hsu, and C. F. Hsu, "A Two-Layer Subband/H261 Video Coder for ATM Networks", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[75] Mohammad Ghanbari and Charles J. Hughes, "Packing Coded Video Signals into ATM Cell", *IEEE/ACM Transactions on Networking*, vol. 1, Oct. 1993.

[76] R. Aravind, R. Civanlar, and Amy R. Reibman, "Cell loss resilience of mpeg-2 scalable video coding algorithms", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[77] Martyn J. Riley and Iain E. G. Richardson, "Reducing Cell Loss Effects on MPEG Video by Varying Slice Size", *in 7th Workshop in Packet Video*, Brisbane, Australia, Mar. 1996.

[78] Matthew Leditschke and Michael Biggar, "Slice Based Approaches to Improving the Error Robustness of MPEG-like Compressed Video Bitstreams", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[79] Pramod Pancha and Magda El Zarki, "MPEG Coding for Variable Bit Rate Video Transmission", *IEEE Communications Magazine*, vol. 32, pp. 54–66, May 1994.

[80] Wenjun Luo and Magda El Zarki, "Adaptive Data Partitioning for MPEG-2 Video Transmission over ATM Based Networks", 1995.

[81] Iain E. G. Richardson and Martyn J. Riley, "Intelligent Packetising of MPEG Video Data", *in Proc. SPIE Visual Communications and Image Processing*, Taipei, May 1995.

[82] Shu Lin and Daniel Costello Jr, *Error Control Coding: Fundamentals and Applications*, Computer Applications in Electrical Engineering. Prentice-Hall, 1983.

[83] A.J. McAuley, "Reliable Broadband Communications Using a Burst Erasure Correcting Code", *in ACM SIGCOMM'90*, pp. 297–306, Philadelphia, PA, 1990. ACM, Academic Press.

[84] Ernst W. Biersack, "Performance Evaluation of Forward Error Correction in an ATM Environment", *IEEE Journal on Selected Areas in Communications*, vol. 11, 1994.

[85] Masahisa Kawashima, "Analysis on the effectiveness of forward error correction code in vbr video transmission", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[86] Yasser Rasheed and Alberto Leon-Garcia, "AAL1 with FEC for the Transport of CBR MPEG2 Video over ATM Networks", *in INFOCOM*, San Francisco, USA, 1996.

[87] Yoshiaki Kato, Kazuhiro Matsuzaki, Tokumichi Murakami, and Hideo Ohira, "A study of fec methods for the atm transmission of mpeg-2 transport stream packets", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[88] Hidenobu Harasaki and Mitsuharu Yano, "Interleaved Forward Error Correction for Variable Bit Rate Video Coding", *in GLOBECOM*, 1994.

[89] Helena Wannman, "Need for cell loss and bit error detection and correction in AAL for AMS", Technical Report af94-0836, Telia Research, Sep. 1994.

[90] Mike Nilsson, David Beaumont, and Geoff Morrison, "An AAL for class B services", Technical Report af94-0780, British Telecom, Sep. 1994.

[91] Naotaka Morita and Masahisa Kawashima, "Review of the Existing Cell Interleave Method for Realtime Data Transport and Proposed Scope and Work Items for FEC over ATM", Technical Report af95-1216, NTT Network Switching System Laboratories, Oct. 1995.

[92] V. Parthasarathy, J. W. Modestino, and K. S. Vastola, "Design of a transport protocol for variable-rate video transmission over atm networks", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[93] J. W. Modestino V. Parthasarathy and K. S. Vastola, "Source and Channel Coding Issues for ATM Networks", *in 7th Tyrrhenian International Workshop on Digital Communications*, pp. 124–134, Viaggio, Italy, Sep. 1995.

[94] Hiroshi Ohta and Tokuhiro Kitami, "A Cell Loss Recovery Method Using FEC in ATM Networks", *IEEE Journal on Selected Areas in Communications*, vol. 9, Dec. 1991.

[95] M. Eyvazkhani and B. Nabati, "A New Error Control Scheme for Video Signal Transmission over ATM Network", *in ATM Networks and LANs*, 1996.

[96] Georg Carle, "Towards Scaleable Error Control for Reliable Multipoint Services in ATM Networks", *in Proceedings of the Twelve International Conference on Computer Communications*, Seoul, Korea, Aug. 1995.

[97] Ji Zhang and Mike Perkins et al., "Carrying MPEG-2 Transport Streams Using AAL-5", Technical Report af94-0879, DiviCom, Sep. 1994.

[98] Keiji Tsunoda, Kumiko Kanai, and Hiroshi Esaki, "A Reed-Solomon Erasure Code and its Applications to AAL", Technical Report af95-1162, Toshiba Corporation, Oct. 1995.

[99] Hiroshi Esaki, Kumiko Kanai, and Yoshiyuki Tsuda, "Necessity of Cell Level FEC Policy for Data Transmission Service", Technical Report af95-0150, Toshiba Corporation, Feb. 1995.

[100] Hiroshi Esaki, Georg Carle, Tim Dwight, Aloke Guha, Keiji Tsunoda, Kumiko Kanai, Komukai Toshiba-cho, and Saiwai Ku, "Necessity of an FEC Scheme for ATM Networks", Technical Report af95-0325R2, Toshiba Corporation and University of Karlsruhe and MCI Telecommunications Corp and Network Systems Corporation and Kawasaki, Oct. 1995.

[101] Aloke Guha, Hiroshi Esaki, Georg Carle, and Tim Dwight, "Necessity of Cell-Level FEC Scheme for ATM Networks", Technical Report af95-1011, Network Systems Corp. and Toshiba Corp. and University of Karlsruhe and MCI Telecommunications Corp., June 1995.

[102] Hiroshi Esaki, Georg Carle, and Tim Dwight, "Benefits of AAL-Level FEC Scheme for ATM Networks", Technical Report af95-1437, Toshiba Corp. and University of Karlsruhe and MCI Telecommunications Corp., Dec. 1995.

[103] Hiroshi Esaki, Georg Carle, and Tim Dwight, "Reliable Multicast Service Needs Cell-Level FEC Scheme", Technical Report af95-1438, Toshiba Corp. and University of Karlsruhe and MCI Telecommunications Corp., Dec. 1995.

[104] Horishi Esaki, Georg Carle, Tim Dwight, Aloke Guha, Keiji Tsunoda, and Kumiko Kanai, "Draft Proposal for Specification of FEC-SSCS for AAL type 5", Technical Report af95-0326R2, Toshiba Corp. and University of Karlsruhe and MCI Telecommunications Corp. and Network Systems Corp., Oct. 1995.

[105] G. Ratta, Mark W. Garret, Gray Hayward, Tony McAuley, and Teun Ott, "On the Proper Place for FEC", Technical Report af95-154, Bellcore, Dec. 1995.

[106] Georg Carle, Aloke Guha, Tim Dwight, Kei-ji Tsunoda, Kumiko Kanai, and Hiroshi Esaki, "A Cell-Level Forward Error Correction Scheme (FEC-SSCS) for ATM Networks", *IEEE/ACM Transactions on Networking*, vol. , Aug. 1995.

[107] Martyn J. Riley and Iain E. G. Richardson, "An FEC Protocol for MPEG2 Video over ATM", to be published, Oct. 1996.

[108] K. Jeffay, D.L. Stone, and F.D. Smith, "Transport and Display Mechanisms for Multimedia Conferencing Across Packet-Switched Networks", *Computer Networks and ISDN Systems*, vol. 26, pp. 1281–1304, 1994.

[109] Oivind Kure and Ingvild Sorteberg, "XTP over ATM", *Computer Networks and ISDN Systems*, vol. 26, pp. 253–262, 1993.

[110] H. Dietmar R. G. Herrtwitch O. Krone J. Sandvoss L. Delgrossi, C. Halstrick and C. Vogt, "Media Scaling with HeiTS", Technical report, IBM European Networking Center, 1994.

[111] C. Topolcic, "Internet Stream Protocol Version 2 (ST-II)", Technical Report RFC 1190, Internet Engineering Task Force, Oct. 1990.

[112] L. Delgrossi, C. Halstrick, R. G. Herrtwitch and H. Stüttgen, "HeiTP - A Transport Protocol for ST-II", *in GLOBECOM'92*, Orlando, USA, 1992.

[113] Luca Delgrossi, *Design of Reservation Protocols for Multimedia Communications*, Kluwer Academic Publishers, 1996.

[114] Rainer Schatzmayr and Radu Popescu-Zeletin, "Providing support for data transfer in a new networking environment", *in Multimedia Transports and Teleservices, International COST 237 Workshop*, Vienna, Austria, 1994.

[115] Gunnar Karlsson, "ATM Adaptation for Video", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[116] L. Delgrossi and L. Berger, "Internet Stream Protocol Version 2 (ST2) Protocol Specification - Version ST2+", Technical Report RFC 1819, Internet Engineering Task Force, 1995.

[117] M.T. Andrade and A.P. Alves, "Functional Architecture if an ATM Channel Adapter for MPEG-2 Applications", *in Sixth International Workshop on Packet Video*, Portland, USA, Sep. 1994.

[118] Teresa Andrade and A. Pimenta Alves, "Transmission of MPEG2 Applications over ATM Networks", *in Multimedia Transports and Teleservices, International COST 237 Workshop*, Vienna, Austria, 1994.

[119] Henry V. Todd and James S. Meditch, "Encapsulation Protocols for MPEG Video in ATM Networks", *in INFOCOM*, San Francisco, 1996.

[120] Carlos Lopez Fernandez, Andrea Basso, and Jean-Pierre Hubaux, "Error concealment and early resynchronization techniques for mpeg-2 video streams damaged by transmission over atm networks", *in SPIE*, San Jose, USA, February 1996.

[121] Carlos Lopez Fernandez, Andrea Basso, and Jean-Pierre Hubaux, "Recovery of DC Levels in Early Resynchronized MPEG-2 Video Streams Damaged by ATM Transmission Errors", *in 7th Workshop in Packet Video*, Brisbane, Australia, Mar. 1996.

[122] J. Youn S. Lee and S. Jang, "Transmission error detection, resynchronization, and error concealment for mpeg-2 video decoder", *in SPIE*, 1993.

[123] Simon Znaty, Thomas Walter, Marcus Brunner, Jean-Pierre Hubaux, and Bernhard Plattner, "Multimedia Multipoint Teleteaching over the European ATM Pilot", *in IZS'96*, pp. 225–238, Zurich, 1996.

[124] Anthony Alles, "ATM Internetworking", Technical report, Cisco Systems, 1995.

[125] Ralf Guido Herrtwitch Luca Delgrossi and Frank Oliver Hoffmann, "An Implementation of ST-II for the Heidelberg Transport System", *Internetworking: Research and Experience*, vol. 5, 1994.

[126] E. Gauthier, J.-Y. Le Boudec, and P. Oechslin, "Shared many-to-many reservations", *in IEEE ATM'96 Workshop*, San Francisco, USA, Aug. 1996.

[127] ITU-T Study Group 13, editor, *Recommendation I.363.5 B-ISDN ATM Adaptation Layer Specification, Type 5*, ITU-T, Geneva, July 1995, Draft.

[128] ITU-T, editor, *H.222.1 Multimedia Multiplex and Synchronization for Audiovisual Communications*, ITU-T, Geneva, 1995, Draft.

[129] Paolo Coppo, Mario Ullio, Lars Lundgren, and Loris Gummerus, "Service Requirements for Multimedia Desktop", Technical Report af94-0355, CSELT and Ericsson Business Networks AB, May 1994.

[130] Paolo Coppo, Mario Ullio, Lars Lundgren, and Loris Gummerus, "Service Requirements for Video Conferencing", Technical Report af94-0356, CSELT and Ericsson Business Networks AB, May 1994.

[131] Charles van Dusen and Mayer Schwartz, "Service Requirements for Interactive Distance Learning", Technical Report af94-0469, Tektronix, Inc, May 1994.

[132] Ron Wiles and Sylvia Wofford, "Service Requirements for Video-on-Demand ATM Transport", Technical Report af94-0688, US WEST Technologies and Ericsson Network Systems, July 1994.

[133] Guido Franceschini, "Support to VBR Real Time Services in the AAL", Technical Report af94-0636, CSELT, July 1994.

[134] Gary Longston and Tony Wasilewski, "MPEG in ATM Networks and a Proposal for a VBR Video Adaptation Layer: AAL6", Technical Report af93-xxx, Scientific Atlanta, Jan. 1994.

[135] Methi Methiwalla, "Video Requirements and the Jitter Issue", Technical Report af94-0472, Northern Telecom, May 1994.

[136] Paul Hodgins and Eisaburo Itakura, "The Issues if Transportation of MPEG over ATM", Technical Report af94-0570, Research Center Information and Communication, Sony Corporation, July 1994.

[137] Ashraf S. Patel, "Outline for the SAA AMS Basline Text Document", Technical Report af94-0279R1, Siemens Stromberg-Carlson, May 1994.

[138] Guido Franceschini, "Extension of the Adaptable Clock Method to Variable Bit Rate Streams", Technical Report af94-0321, CSELT, Apr. 1994.

[139] Ron Wiles, "CBR MPEG over ATM", Technical Report af94-0343, US WEST Technologies, May 1994.

[140] Ted Brunner and Mayer Schwartz, "Essential Features of a Video and Audio SSCS for AAL5", Technical Report af94-0404, Tektronix, Inc, May 1994.

[141] Jeff Lynch et al. Dave Davenport, "Video Distribution using CBR MPEG-2 over AAL5", Technical Report af94-0857, IBM Corporation and Sony Corporation, Sep. 1994.

[142] Paul Skelly, "Proposal for MPEG-2 Transport over ATM", Technical Report af94-0842, Bellcore, Sep. 1994.

[143] Ron Wiles, "AMS Implementation Agreement Version 1.0", Technical Report af94-0896, US WEST Technologies, Sep. 1994.

[144] David Beaumont, "Report on European Progress on AALs for Audio-visual and Multimedia Services", Technical Report af94-1151, British Telecom, Nov. 1994.

[145] Paul Hodgins and Eisaburo Itakura, "AAL5 Encapsulation of MPEG-2", Technical Report af94-1129, Research Center Information Division and Communication, Sony Corporation, Nov. 1994.

[146] Sudhir S. Dixit and Ashok Kumar, "Proposals for Carrying CBR MPEG-2 Transport Streams over ATM for Interactive and Broadcast Video-Audio Services", Technical Report af94-1124, NYNEX and RAYNET Corporation, Nov. 1994.

[147] Mike Nilsson and David Beaumont, "Proposal for AAL-5 Packing Scheme for the Audio-visual Multimedia Services Implementation Agreement", Technical Report af95-1150, British Telecom, Oct. 1995.

[148] Paul Hodgins, "More on MPEG-2/AAL5 Encapsulation", Technical Report af95-1125, Sony Telecom Europe, Oct. 1995.

[149] Jeff Lynch, "Encapsulating MPEG-2 TS Packets into AAL5 PDUs", Technical Report af94-1000, IBM Corporation, Dec. 1994.

[150] David Singer and Mengjou Lin, "MPEG2/AAL5 Proposed PDU sizing", Technical Report af94-1047, Apple Computer, Nov. 1994.

[151] Sudhir S. Dixit, "Aal, timing recovery and error detection/correction for mpeg over atm", Technical Report af94-0100, NYNEX, Jan. 1994.

[152] Christian Wipliez, "SAA-AMS: France Telecom Comments on the Packing Issue", Technical Report af95-1138, France Telecom, Oct. 1995.

[153] Christian Wipliez, "SAA-AMS: Experimental Measurements on ATM Networks QOS: A Guideline for the Choice of AALs", Technical Report af95-1246, France Telecom CNET, Oct. 1995.

[154] Timothy W. Dygert and Kenneth G. Stocker, "MPEG AAL Packing Perspective - Justification for Reexamining the Choice of AAL5 versus AAL1 for ATM Transport of MPEG", Technical Report af95-1322, Stellar One Corporation, Oct. 1995.

[155] The ATM Forum Technical Committee, *Video on Demand Specification 1.1*, Straw vote edition, Oct. 1996.

[156] Jeff Lynch, "Ams phase 2 living list", Technical Report af96-0671R2, IBM Corporation, Oct. 1996.

[157] Members of the Applications Technical Committee of DAVIC, *DAVIC 1.0 Specification, Part 7: Mid-Layer Protocols*, Digital Audio-Visual Council, 1995.

[158] D. Clark and D. Tennenhouse, "Architectural considerations for a new generation of protocols", *in ACM SIGCOMM'90*, pp. 201–208, Sep. 1990.

[159] Sally Floyd, Van Jacobson, and Steve McCanne, "A reliable multicast framework for light-weight sessions and application level framing", *in ACM SIGCOMM'95*, pp. 342–356, Aug. 1995.

[160] Donn Hoffman, Gerard Fernando, and Vivek Goyal, "RTP Payload Format for MPEG1/MPEG2 Video", Technical Report draft-ietf-avt-mpeg-02.txt, Internet Engineering Task Force, 1996, Expires on February 1, 1997.

[161] The ATM Forum Technical Committee, *Video on Demand Specification 1.0*, First edition, Jan. 1996.

[162] Bernd Lamparter, Otto Böhrer, Wolfgang Effelsberg, and Volker Turau, "Adaptable Forward Error Correction for Multimedia Data Streams", Technical report, University of Mannheim, Sep. 1993.

[163] Thomas Stock and Xavier Garcia Adanez, "On the Potentials of Forward Error Correction Mechanisms Applied to Real-Time Services Carried over B-ISDN", *in IZS'96*, pp. 107–118, Zurich, 1996, available on `http://tcomwww.epfl.ch/~garcia`.

[164] Xavier Garcia Adanez, "Opnet user guide", Technical report, Telecommunications Laboratory, Swiss Federal Institute of Technology, 1993.

[165] Chadd Fogg, "mpeg2encode/mpeg2decode version 1.1. available via anonymous ftp at ftp.netcom.com", *in MPEG Software Simulation Group*, June 1994.

[166] Olivier Verscheure, "Perceptual activity metric", *in SPIE 96*, San Jose, USA, 1996.

[167] ITU-T, editor, *Recommendation H.310 Broadband and Audiovisual Communications Systems and Terminals*, ITU-T, Geneva, Nov. 1996.

[168] Maher Hamdi, *Contrôle de Traffic ATM pour Sources Vidéo à Débit Variable*, PhD thesis, Département Réseaux et Services Multimedia, Télécom Breatgne, Université de Rennes, 1996.

[169] Rolf Keller, "Implantation d'une couche de protocole AAL pour applications multimédia sur ATM", Master's thesis, Laboratoire de Télécommunications, Ecole Polytechnique Fédérale de Lausanne, 1996.

Index

Xavier Garcia Adanez

School address
EPFL/DI-ICA
CH-1015 Lausanne
Switzerland
Voice: (+41 21) 693-5258
Fax: (+41 21) 693-2683
EMail: xavier.garcia@epfl.ch
URL: http://tcomwww.epfl.ch/~garcia

Personal address
4, Petit-Beaulieu
1004 Lausanne
Switzerland
(+41 21) 646-1294

Education

Swiss Federal Institute of Technology — **Lausanne, Switzerland**
Department of Electrical Engineering.
Ph.D. program in Communication Systems.
Mar. 1992 - present.

Swiss Federal Institute of Technology — **Lausanne, Switzerland**
Department of Electrical Engineering.
Diploma in Electrical Engineering (M.Sc.).
Swiss Federal Federal Institute of Technology.
Oct. 1985 - Jan. 1992.

Lycée Français Jean Mermoz — **Buenos Aires, Argentina**
Baccalauréat Français de type C (High School Degree)
November 1984.

Experience

Sep. 97 - present	**Swiss Federal Institute of Technology** **Institute for computer Communications and Applications** Research Assistant and Ph.D. Candidate.
Mar. 92 - Sep. 97	**Telecommunications Laboratory** Research Assistant and Ph.D. Candidate.

Projects

INCOMS	Interworking Concepts for Multimedia Systems over B-ISDN. Joint project with Alcatel STR.
Ernestine II	Performance studies of the SS7 network for the development of new services on the intelligent network. Joint Project with Swiss Telecom PTT

Languages

Spanish, French	native languages
English	fluent
Italian	adequate knowledge

Publications

1 Xavier Garcia Adanez and Olivier Verscheure, "Networking from the Application Perspective", COST 237 Final Report, Chapter1, in preparation.

2 Xavier Garcia Adanez and Olivier Verscheure, "New Network and ATM Adaptation Layers for Interactive MPEG-2 Video Communications: A Performance Study Based on Psychophysics", accepted for publication in Interoperable Communication Networks Journal, Baltzer Science Publisher, 1997.

3 Xavier Garcia Adanez, Olivier Verscheure and Pascal Frossard, "Reliable Transmission of MPEG-2 VBR Video Streams over New Network Adaptation and ATM Adaptation Layers", IEEE ATM'97 Workshop, Lisbon Portugal, May 1997.

4 Xavier Garcia Adanez, Olivier Verscheure and Jean-Pierre Hubaux, "New Network and ATM Adaptation Layers for Real-Time Multimedia Applications: A Performance Study Based on Psychophysics", in Multimedia Transports and Teleservices, International COST 237 Workshop, Barcelona, Spain, Nov 1996.

5 Olivier Verscheure and Xavier Garcia Adanez, "Perceptual Quality Metric as a Performance Tool for ATM Adaptation of MPEG-2 based Multimedia Applications", in EUNICE Summer School on Telecommunications Services, Swiss Federal Institute of Technology, Lausanne, Sep. 1996.

6 Xavier Garcia Adanez et Olivier Verscheure, "Application d'une Mesure de Qualité Vidéo Perceptive à l'Analyse des Performances de AAL-5 et d'une Nouvelle AAL sur des Flux Vidéo MPEG-2", in Journées Doctorales en Informatique et Réseau, JDIR'96, Paris, Sep. 1996.

7 Xavier Garcia Adanez and Andrea Basso and Jean-Pierre Hubaux, "Study of AAL5 and a New AAL Segmentation Mechanism for MPEG-2 Video over ATM", 7th Workshop in Packet Video, Brisbane, Australia, Mar. 1996.

8 Thomas Stock and Xavier Garcia Adanez, "On the Potentials of Forward Error Correction Mechanisms Applied to Real-Time Services Carried over B-ISDN", International Zurich Seminar, Springer, Feb. 1996, pp. 107-118.

8 Sam Manthorpe and Xavier Garcia Adanez, "TCP performance over ATM Based LAN Interconnection Services", Proc. Interop Engineering Conference, Las Vegas, May 1995.

9 Sam Manthorpe and Xavier Garcia Adanez, "Shaper Dimensioning for MAN to ATM Based WAN", Proc. Interop Engineering Conference, Las Vegas, May 1994.